JOIM
www.joim.com

JOURNAL OF INVESTMENT MANAGEMENT, Vol. 15, No. 3, (2017), pp. 23–40

MARKET TIMING: SIN A LITTLE RESOLVING THE VALUATION TIMING PUZZLE

Cliff Asness[a], *Antti Ilmanen*[b] *and Thomas Maloney*[c]

Successful market timing is a tantalizing holy grail for investors, especially when there seems to be persuasive evidence that simple valuation measures can predict subsequent market performance. But, as both researchers and investors have discovered, outperforming a passive buy-and-hold approach is harder than it might seem. Is market timing a useful source of added value or a sin to be avoided? In this paper we explore the difference between the encouraging in-sample long-horizon evidence and directionally right but weak and disappointing out-of-sample performance. We propose an interpretation that offers a practical enhancement to value timing strategies: adding a dose of momentum.

For the faction of investors open to or enthusiastic about the idea of market timing, especially those with long horizons, contrarian strategies are by far the most popular and well-regarded. This preference, which stems from both economic intuition[1] and evidence of long-horizon return predictability, may be too hasty.

In light of the statistical evidence that valuation ratios forecast subsequent long-horizon returns, this paper examines the performance of market timing strategies based on these same valuation ratios. It finds that while contrarian market timing has outperformed buy-and-hold over the past 115 years, it has underperformed in the latter half of the sample (a very long time!) and generally looks weaker than many might expect. This paper explores why, despite seemingly strong in-sample statistical evidence, contrarian market timing has underperformed over the past six decades. First, we show that drifting valuations have made recent decades a particularly challenging period for contrarian timing that is unlikely to be repeated. Second and more importantly, we show that contrarian timing strategies are fighting the successful shorter-term momentum strategy.

[a]Cliff Asness is founding and managing principal at AQR Capital Management, Two Greenwich Plaza, Greenwich, CT 06830, E-mail: cliff.asness@aqr.com
[b]Antti Ilmanen is a principal at AQR Capital Management Europe, Charles House, 5-11 Regent Street, London SW1Y 4LR, UK. E-mail: antti.ilmanen@aqr.com
[c]Thomas Maloney is a vice president at AQR Capital Management Europe, Charles House, 5-11 Regent Street, London SW1Y 4LR, UK. E-mail: thomas.maloney@aqr.com

We propose that by adding a dose of momentum, one can improve the risk-adjusted performance of contrarian market timing strategies. In fact, a naïve reading of the historical experience suggests using mostly or even only momentum-based timing, but diversification logic and more careful empirical analysis supports combining both contrarian and momentum indicators to make modest market timing tilts.

1 The valuation timing puzzle

The subject of equity market timing has attracted an extensive literature, which has tended to agree[2] on the existence of a link between some measures of stock market valuations and some measures of subsequent returns. The literature has, however, disagreed on both the best method of assessing this relationship—statistical measures or benchmark performance comparisons—and on the practical implications for investors seeking to beat a buy-and-hold approach out of sample.

Goyal and Welch (2008) reexamine many prominent timing indicators from earlier studies, focusing on statistical forecasting power rather than outperformance versus a benchmark. They conclude that evidence for robust out-of-sample forecasting of the equity premium is almost uniformly unconvincing. They reiterate the important observation that selection and data-mining biases are endemic in market timing studies. Dimson *et al.* (2013) come to similar conclusions, showing first promising in-sample patterns, but then much weaker out-of-sample results and an inability to outperform 100% equities by switching to cash when valuations are unattractive.[3]

But Campbell and Thompson (2008) argue that simple, economically intuitive restrictions on the regressions of Goyal and Welch give much stronger performance of predictive indicators. They also show that even a small amount of predictive power can be economically meaningful for investors.[4]

Comparisons of predictive power at different horizons must be made carefully. A series of longer-horizon returns with built-in persistence will mechanically tend to produce higher correlations, and even with a century of data the number of independent long-horizon observations is small. Even so, evidence of a link between valuation and subsequent returns does generally appear stronger for longer horizons (next 5–10 years' return) than for shorter horizons (next quarter). We show that short-term momentum may help to explain this result.

We begin by summarizing the oft-cited statistical evidence, and then put it to the test in a realistic out-of-sample U.S. equity strategy with contrarian tilts. In the Appendix we show a similar analysis for U.S. Treasuries, with similar results.

2 Unpacking the statistical evidence

Chart 1 shows the average rate of excess return (over cash) for U.S. equities for 10-year periods measured from the start of each quarter and sorted by starting valuation, as measured by the Shiller P/E or CAPE ratio,[5] using over a century of data from January 1900 to December 2015. The evidence for higher valuations predicting lower subsequent returns (and vice versa) does indeed appear strong.[6]

There are several reasons to suspect that real-life market timing strategies will not deliver as strong results as Chart 1 seems to promise. First, the chart involves an important hindsight bias: we define the quintiles using the full history. In other words, each quarter we evaluate the market relative to both past and future valuations. Real-time investors do not know how future valuations may evolve and change the definition of what constitutes high or low valuations.

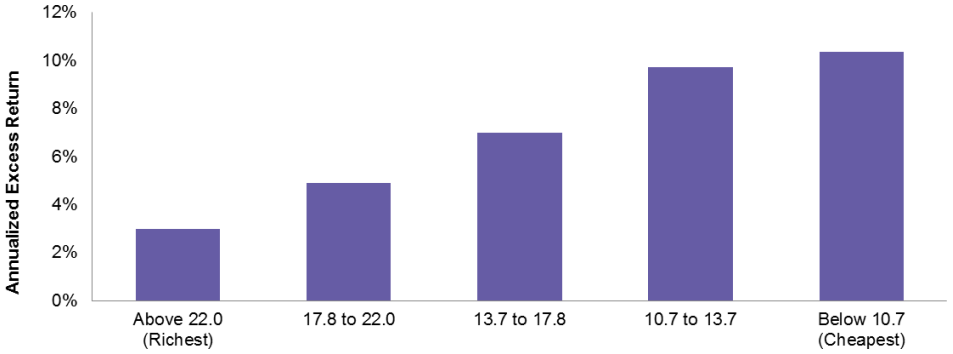

Chart 1 U.S. Equity 10-year returns sorted by starting CAPE Valuation, 1900–2015.

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. U.S. equity here and henceforth is the S&P 500 Index from 1926, and prior to 1926 a reconstruction of the S&P 500 available on Robert Shiller's website which uses dividends and earnings data from Cowles and Associates, interpolated from annual data. Annualized arithmetic mean rates of return excess of cash, based on quarterly data.

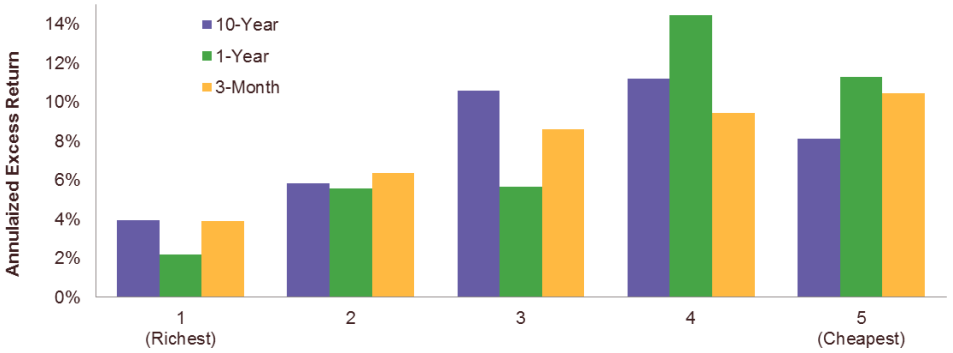

Chart 2 U.S. equity returns sorted by starting valuation based on rolling 60-year window, 1900–2015.

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Annualized arithmetic mean rates of return excess of cash, based on quarterly data.

Chart 2 removes this bias by defining quintiles using a rolling 60-year window of past data.[7] It also adds 1-year and 3-month returns, to see if the pattern holds at shorter horizons. With an out-of-sample approach the patterns are weaker, particularly in the disappointing fifth quintile, but they are still visible.

More formally, we can test for explanatory power using OLS regressions. Table 1 shows the results of in-sample regressions that appear encouraging (statistically significant at both long and short horizons) but suffer from a similar hindsight bias to the pattern in Chart 1. We also tested for out-of-sample (OOS) predictive power using similar methods as earlier researchers. Results were generally weaker, as we would expect, but were also highly dependent on the method and the choice of parameters.[8]

Table 1 Shiller EP as explanatory variable of future equity returns 1900–2015 (in-sample).

	Next 10Y	Next 1Y	Next 1M
Beta	0.8	1.3	1.2
T-statistic	3.4	1.9	2.3
R-squared	0.24	0.05	0.00

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Single-factor regressions based on annualized monthly returns. Newey–West-adjusted *T*-stats accounting for overlapping observations.

3 Contrarian tilts vs. buy-and-hold

While the above evidence suggests that valuation is a useful contrarian market timing signal, it is not a test of a realistic trading strategy. We construct a simple tactical timing strategy that scales its equity investment in the range 50–150% (effectively adding a tactical overlay to a buy-and-hold portfolio), and then compare this strategy with a fully-invested buy-and-hold approach. This method of appraisal has the advantage of comparing passive and active portfolios with equal expected average exposure to the equity premium, and avoids the tendency of other studies to apply extreme, one-sided tilts (e.g., hold either stocks or cash).[9]

Specifically, the timing strategy applies a weight of 100% + (trimmed Shiller *EP*−median Shiller *EP*)/(95th–5th percentile range), with a floor at 50% and a cap at 150%. The trimming (at 95th and 5th percentiles) helps to reduce the distorting or compressing effects of extreme values. This signal function was designed to be reasonable *ex ante*, rather than selected based on results, and indeed most simple or reasonable methods produce similar results (see Appendix).[10] We use data from 1881 with an expanding percentile window until 1941 and then a rolling 60-year window, as we did for Chart 2. We rebalance monthly, borrowing or lending cash with the rest of the portfolio.

Before discussing the results, we must decide on our evaluation criteria. Return is the most tangible, but we also consider risk-adjusted returns to determine whether any increases in risk from market timing are adequately compensated. Table 2 shows basic performance statistics, whereas Chart 3 shows cumulative returns and the underlying signal. The results are disappointing. Even before costs, the timing strategy cannot beat the Sharpe ratio[11] of buy-and-hold over either the full 116-year sample or the latter half of it (this starts in 1958). During this latter period, it has earned lower returns than buy-and-hold.[12] Neither return differences (difference of means test) or Sharpe ratio differences (Jobson–Korkie test) are statistically significant over either period.

This is a puzzle. Given the strong in-sample evidence, why are these results so weak? The most important reason for the lower returns in recent decades is highlighted in the last two rows of

Table 2 Hypothetical performance of buy-and-hold and simple timing strategies in U.S. equities, 1900–2015.

	1900–2015		1958–2015	
	Buy-and-hold	Value timing	Buy-and-hold	Value timing
Excess return	6.6%	7.4%	5.5%	5.4%
Volatility	17.5%	20.0%	14.9%	14.6%
Sharpe ratio	0.38	0.37	0.37	0.37
Max drawdown	−83%	−87%	−53%	−51%
Max relative drawdown		−32%		−32%
Average position	100%	102%	100%	89%
Impact of avg. tilt		0.1%		−0.6%

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Hypothetical performance excess of cash, gross of *t*-costs and fees, with monthly rebalancing. Arithmetic returns and Sharpe ratios. Drawdowns are based on total returns. 'Max relative drawdown' is maximum cumulative underperformance vs. buy-and-hold.

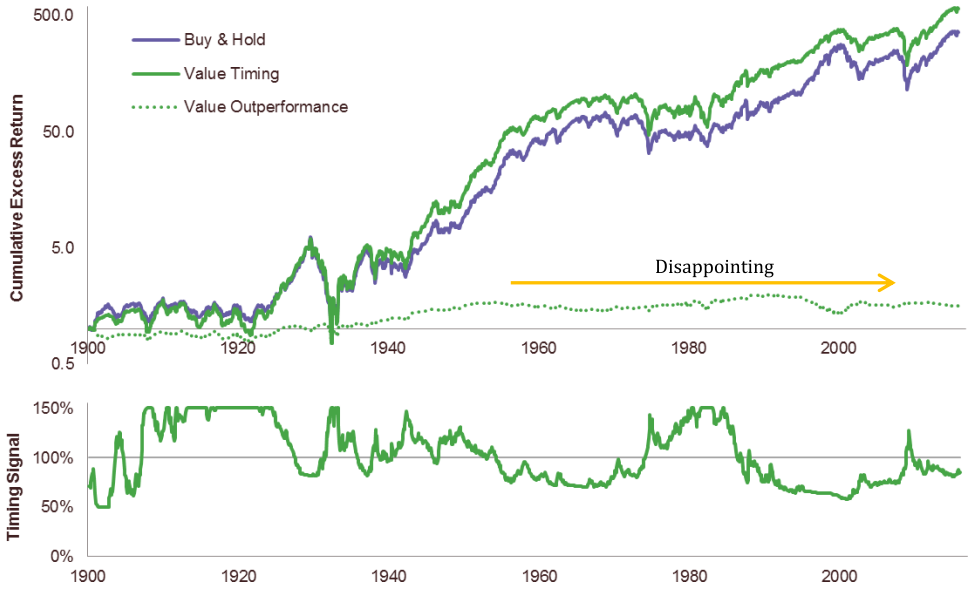

Chart 3 Hypothetical cumulative return of buy-and-hold and simple timing strategies in U.S. equities, 1900–2015.

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Hypothetical performance excess of cash, gross of *t*-costs and fees, with monthly rebalancing.

the table: the timing strategy has been *underinvested* on average (average position 89% since 1958). While the Shiller P/E ratio generally drifted *lower* during the early 1900s, it generally drifted *higher* during the last 60 years (see rolling median in Chart 4). This upward drift means the timing strategy gets a disproportionate number of "underweight" signals in recent decades. If higher-frequency contrarian timing signals had been accurate enough, they might have overcome the return drag of −0.6% per annum from the forfeited equity premium (calculated by multiplying the average timing tilt of 11% by the buy-and-hold annual return over the same period). However, this has not been the case.

Another way to illustrate the effect of this drift in valuations is the "back-expanding median" CAPE shown in Chart 4 (yellow line), which is the median, as of 2015, calculated from each

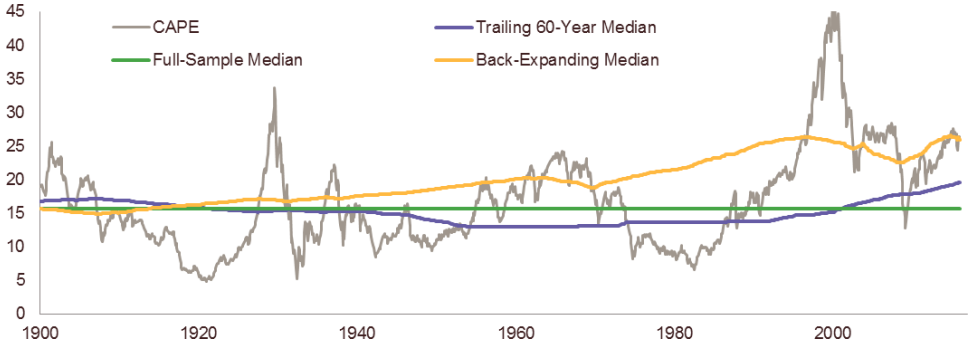

Chart 4 CAPE ratio, U.S. equities 1900–2015.

Source: Robert Shiller's Data Library.

specified start-date to the end of the sample (so, for example, the 2010 data-point is the median over the period 2010–2015, and the 2000 data-point is the median over the period 2000–2015). In our strategy, and most similar strategies, the median CAPE represents the neutral position where no timing tilt is undertaken. The chart illustrates that our current definition of what constitutes this neutral valuation gradually gets lower as we look further back into the past to calibrate our model.

The observed drift is a sample-specific result, and probably a particularly bad draw for the timing strategy, with the CAPE far above any long-term historical average at the end of the sample (see next section). It does not prove that contrarian timing strategies won't work in the future. But it does illustrate a fundamental difficulty faced by such strategies: valuations can drift higher or lower for years or decades, making it difficult to categorize the current market confidently as "cheap" or "expensive" without hindsight calibration, and therefore difficult to profit from such categorizations.[13] One stark illustration of the challenges contrarian investors face is that in the 1990s our timing strategy gets a strong "overvalued" signal (underweights by at least 25%) not in 1999 or even 1996, but in 1991: a painful case of "early equals wrong."[14]

Several additional factors contribute to the disappointing results. The full-sample Sharpe ratio, for example, is reduced by the tendency of cheap valuations to predict not only higher returns but also higher volatility.[15]

There are of course many variants of, and potential enhancements to the simple value strategy we analyze above (some are described in the Appendix). For example, many observers believe that contrarian timing works better when applied only at extreme valuations. This sounds plausible to us as well, but we show in the Appendix that such tweaks do not meaningfully improve results. Regardless of the design details, the main challenge remains: drifting contrarian indicators make it difficult to evaluate the current market in real time and give profitably time-varying exposure to the equity premium.

4 Leveling the playing field?

If our sample is a bad draw for value timing, what might a more neutral draw look like? Table 3

Table 3 Hypothetical performance of value timing strategies with *ex post* adjustments.

	a. Shifted signal				b. Zero net valuation change			
	1900–2015		1958–2015		1900–2015		1958–2015	
	Buy-and-hold	Value timing	Buy-and-hold	Value timing	Buy-and-hold	Value timing	Buy-and-hold	Value timing
Excess return	6.6%	7.3%	5.5%	6.0%	6.7%	8.1%	3.7%	4.3%
Volatility	17.5%	19.7%	14.9%	16.1%	18.5%	21.5%	15.5%	16.3%
Sharpe ratio	0.38	0.37	0.37	0.37	0.36	0.38	0.24	0.27
Average position	100%	100%	100%	100%	100%	106%	100%	97%

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Hypothetical performance excess of cash, gross of transaction costs and fees, based on monthly rebalancing. Arithmetic rates of returns and Sharpe ratios. Panel (b) shows mean statistics for all periods exceeding 20 years with a change in Shiller *EP* of less than 0.1%.

shows results for two approaches to adjusting the results *ex post* (a.k.a. cheating). The left-hand panel shows the impact of simply shifting the signal by adding a constant, to give an average position of 100% over each sample. After this adjustment, the timing strategy earns higher returns than buy-and-hold over both samples, as promised by the in-sample evidence.

The right-hand panel shows mean performance statistics for all periods exceeding 20 years with a net change in Shiller E/P of less than 0.1%: periods with equal starting and ending valuations.[16] Over these periods specifically selected to have no net drift in valuations, value timing again earns higher returns than buy-and-hold. But note that the timing strategy is still underinvested in the more recent sample: even with investment periods selected to have no net richening, the strategy suffers from richening during calibration periods.[17]

While the above *ex post* adjustments benefit the timing strategy on paper, unfortunately no investor can guarantee either an average tilt of zero or a net valuation change of zero over their own investment horizon. Indeed, the tendency of valuations to drift for long periods is precisely the challenge we discussed in the previous section. Is there a more robust and practical way to improve the prospects of market timing? The data offer a clue: we nearly always find better historical results from *momentum* than contrarian timing. A more promising perspective on the disappointing performance of contrarian strategies may therefore be to examine whether they face an uphill battle against shorter-term momentum.

5 Sometimes underinvested, often fighting momentum

Markets have been shown to exhibit trends or time series momentum at multi-month horizons,[18] and we might expect contrarian timing strategies to be fighting against this headwind. Table 4 uses the following OLS regression to attribute the active return earned by our value tilt:

$$(R_{Timing} - R_{B\&H}) = \alpha + \beta_{Mkt} R_{B\&H} + \beta_{Mom} R_{Mom} + \varepsilon$$

The first beta term is the return explained by inadvertent net market exposure (being over- or underinvested on average over the sample), and the second beta term is the return explained by exposure to a simple time series momentum factor that is constructed in exactly the same way as the contrarian tilt, but using past 12-month excess return instead of Shiller E/P.[19] Alpha is the return unexplained by these two factors.

Market exposure contributes positively in the first half of the sample (the value strategy tends to be

Table 4 Attribution of hypothetical value timing active returns in U.S. equities.

	1900–1957 (Active Ret. +1.7%)			1958–2015 (Active Ret. −0.1%)			Full sample (Active Ret. +0.8%)		
	Market	Mom.	Intercept	Market	Mom.	Intercept	Market	Mom.	Intercept
Coefficient	0.2	−0.3	1.0%	−0.1	−0.3	0.5%	0.1	−0.2	0.6%
T-statistic	2.4	−3.5	1.4	−1.8	−1.9	0.9	1.3	−2.8	1.2
Attribution	1.4%	−0.7%	1.1%	−0.5%	−0.1%	0.5%	0.6%	−0.4%	0.6%

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Two-factor regressions based on monthly returns. Newey–West T-stats accounting for overlapping observations. Hypothetical annualized arithmetic returns excess of cash, gross of transaction costs and fees, based on monthly rebalancing.

overweight), and negatively in the second half of the sample. As discussed previously, these outcomes would have been difficult to predict in advance, and there is not much we can do about them apart from recognizing the risk of long-term drifts in valuations and scaling contrarian tilts modestly.

Negative exposure to the momentum factor, on the other hand, detracts from performance in both sub-samples. This is a potentially useful result, especially given independent supportive evidence for time series momentum in many other asset classes. The intercept can be interpreted as the return realized by our value timing strategy in a world without valuation drift and with the negative exposure to momentum neutralized – it is positive in both samples but not statistically significant.

Combining value and momentum has been shown to be very effective in stock selection and cross-sectional strategies,[20] and the combined signal intuitively represents "value with a catalyst," or patient contrarian investing, with a supportive momentum signal potentially reducing the risk of value traps or premature signals. In Table 5 we show the performance of a combined strategy

Table 5 Hypothetical performance of buy-and-hold and simple timing strategies in U.S. equities.

	1900–2015				1958–2015			
	Buy-and-hold	Value timing	Mom timing	VM timing	Buy-and-hold	Value timing	Mom timing	VM timing
Excess return	6.6%	7.4%	8.1%	7.8%	5.5%	5.4%	5.9%	5.7%
Volatility	17.5%	20.0%	17.1%	18.0%	14.9%	14.6%	13.8%	13.8%
Sharpe ratio	0.38	0.37	0.48	0.43	0.37	0.37	0.43	0.41
Average position	100%	102%	102%	102%	100%	89%	97%	93%

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Hypothetical performance excess of cash, gross of *t*-costs and fees, with monthly rebalancing. Arithmetic returns and Sharpe ratios.

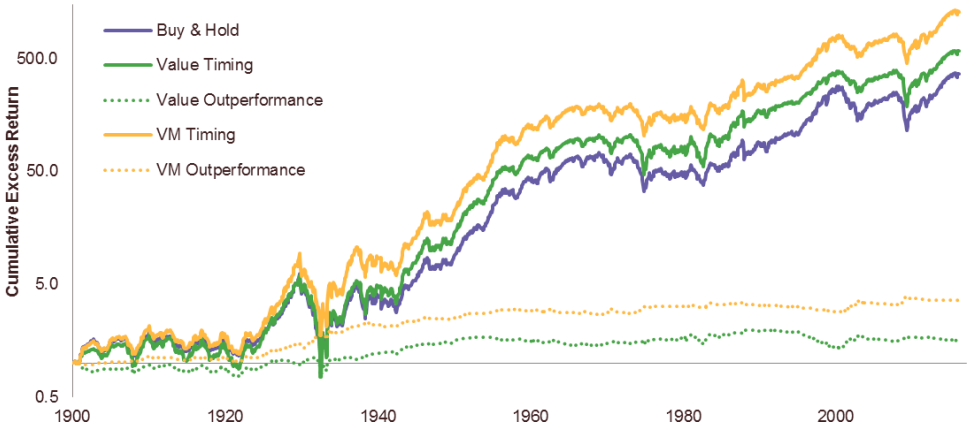

Chart 5 Hypothetical cumulative return of buy-and-hold and simple timing strategies in U.S. equities.

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Hypothetical performance excess of cash, gross of *t*-costs and fees, with monthly rebalancing.

Table 6 Hypothetical worst return outcomes for buy-and-hold and simple timing strategies in U.S. equities.

	1900–2015				1958–2015			
	Buy-and-hold	Value timing	Mom timing	VM timing	Buy-and-hold	Value timing	Mom timing	VM timing
Sharpe ratio	0.38	0.37	0.48	0.43	0.37	0.37	0.43	0.41
Max drawdown	−83%	−87%	−74%	−80%	−53%	−51%	−39%	−43%
1st Percentile drawdown	−69%	−67%	−62%	−64%	−43%	−39%	−31%	−33%
Worst 3Y return	−81%	−85%	−66%	−77%	−43%	−41%	−28%	−31%
1st Percentile 3Y return	−62%	−61%	−43%	−53%	−36%	−26%	−20%	−22%
Max Relative Drawdown	0%	−32%	−31%	−17%	0%	−32%	−27%	−17%

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Hypothetical performance gross of transaction costs and fees, based on monthly rebalancing. Sharpe ratios based on arithmetic returns. Downside risk measures are based on total returns. 'Worst 3Y return' is the lowest 3-year total return. '1st Percentile drawdown' is the largest drawdown excluding the worst 1% of drawdown observations. '1st Percentile 3Y return' is the lowest 3-year total return excluding the worst 1% of 3-year return observations 'Max relative drawdown' is maximum cumulative underperformance vs. buy-and-hold.

("VM Timing") that uses a simple average of value and momentum signals, as defined in the above analysis. Chart 5 shows cumulative returns.

The combined strategy achieves higher gross Sharpe ratios than buy-and-hold or pure value timing over both samples. The value and momentum signals have a usefully diversifying correlation of −0.2. Evidence and intuition both suggest that adding a momentum signal helps to address the challenges of contrarian timing.

While our results suggest that it may be difficult to achieve more than modest improvements to returns and Sharpe ratios by market timing (the margins of outperformance are again statistically insignificant), even gross of costs,[21] downside risks may also be improved. Table 6 shows some downside risk measures for the same strategies. The combined value and momentum timing strategy beats both pure value and buy and hold on all measures, in both samples. If market timing is a sin, perhaps you could—or even *should*—sin a little.

6 Another puzzle

The first puzzle we addressed in this paper was how the inviting picture for contrarian timing in Chart 1 could be reconciled with the rather underwhelming results of practical market timing strategies in Table 2. Table 5 presents another puzzle: Given that pure momentum-based timing outperforms the combined value and momentum strategy, one might wonder, 'Why not 100% momentum?'

There are several reasons why this might be a step too far. First, recent decades have been particularly challenging for value-based timing, and these conditions are unlikely to be repeated as discussed earlier. Second, while momentum has performed well over the long-term, it has suffered periods of sharp underperformance; the combined strategy has experienced milder drawdowns *relative to buy-and-hold* (see Table 6). Elsewhere in this paper we have quoted total risk measures, but many tactical investors care about, and are judged by, relative performance. A reduced risk of substantial relative underperformance

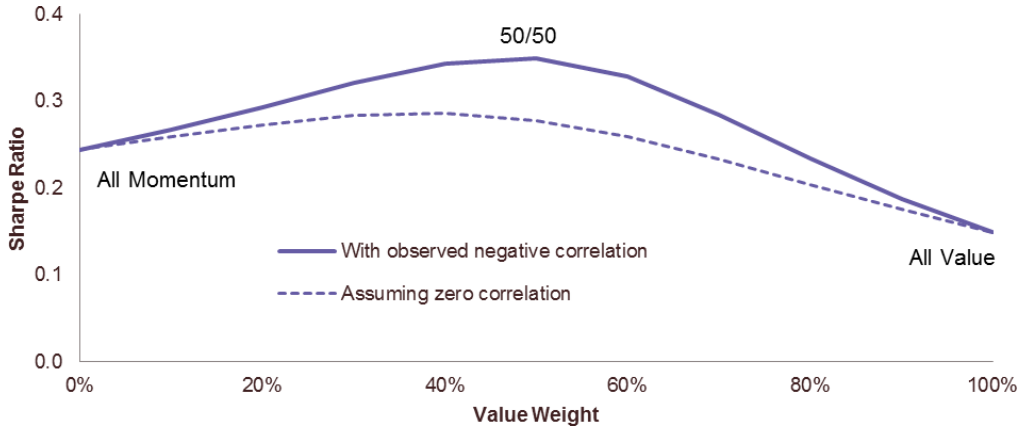

Chart 6 Hypothetical Sharpe ratio of value/momentum timing overlays at different weights, 1900–2015.

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Hypothetical performance excess of cash, gross of *t*-costs and fees, based on arithmetic rates of return and with monthly rebalancing.

may be important to such investors. Third (and related), the negative correlation between value and momentum tilts implies giving some weight to value, even if its standalone performance is disappointing. Finally, momentum has higher turnover and transaction costs, which are not deducted here.[22]

More formally, one can argue that the long-run evidence, together with reasonable priors supports well-balanced VM timing. In a forward-looking analysis, it seems reasonable to assume no correlation between either value or momentum and the market, nor any net market-directional tilt for either style. Then we can study value and momentum timing as standalone long/short strategies, setting aside the question of how and with what weight to combine them with a passive market portfolio.[23] These long/short strategies had standalone Sharpe ratios of 0.15 and 0.24 since 1900, respectively (lower than the 0.37 and 0.48 in Table 5 because these do not include the equity market premium that comes from adding the long/short tilts to a 100% passive position). How would an investor allocate between the two components, *if* they wanted to ignore in-sample interactions with the market (such as the historical tendency of value tilts to be mildly positively correlated to the market), and if they wanted to maximize the risk-adjusted performance (information ratio) of the tactical tilt?

Chart 6 shows the Sharpe ratios of different weighted combinations of the long/short timing strategies. Even though momentum has a markedly higher standalone Sharpe ratio than value, the optimal (i.e., Sharpe ratio maximizing) combination is close to 50/50 (solid line). This is due to the complementary nature of the two strategies (the signals are −0.2 correlated and their single-style timing returns are −0.4 correlated[24]). If the returns correlation was merely zero, rather than negative, the optimal weighting would more clearly favor momentum given its higher Sharpe ratio, and the benefit from combining the strategies would be less (dashed line).

7 Other ways to be a patient contrarian

Value investors looking for a less explicit way to reduce negative exposure to short-term momentum could simply rebalance their tilts less frequently. A better variant of this approach is to "slow down" the value signal using a moving average, which is equivalent to making a sequence of overlapping value bets that are

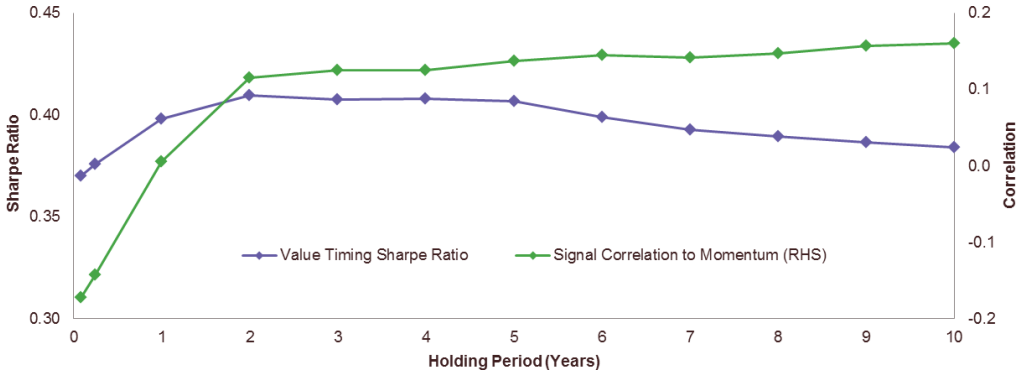

Chart 7 Impact of holding period on hypothetical value timing performance and correlation to momentum, 1900–2015.

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Hypothetical performance gross of transaction costs and fees. Sharpe ratios based on arithmetic returns. Effective holding period increased by smoothing value signal, equivalent to monthly overlapping holding periods.

each "locked in" for a longer period. This directly harnesses the aforementioned stronger relationship between valuation and returns at longer horizons.[25]

Chart 7 shows the full-period Sharpe ratio for our simple value timing strategy with monthly overlapping holding periods of different lengths (purple line, simple monthly strategy at the far left). This is equivalent to smoothing the signal to different degrees. The green line shows the correlation of the resulting net value signal to the momentum factor. The best performance is achieved by forcing a holding period of 2–5 years, which neutralizes the negative exposure to momentum suffered (especially) by the monthly and quarterly strategies. At longer holding periods of 5–10 years performance begins to fall, possibly as valuation signals become stale.

Chart 7 suggests that momentum effects may help explain why valuation is a better predictor at longer horizons than at shorter horizons. Note that while slowing down the value signal (forcing a longer holding period) may improve the Sharpe ratio, it does not address the risk of being under- (or over-)invested due to drifting valuations. By contrast, explicitly adding a diversifying momentum signal can help to dilute inadvertent persistent value tilts, as shown in Table 5.[26] Of course, if, consistent with other research, momentum is the real reason more patient approaches outperform, we would expect these approximations to be somewhat inferior to explicitly accounting for momentum.

8 Implications for the current environment

If we extend our dataset to December 2016, just prior to the time of writing, we find U.S. equities to be somewhat expensive compared to the past 60 years and to have near-neutral momentum, so that the combined signal is a small underweight.[27] A worsening of momentum at or near this valuation would give a particularly bearish combined signal.

There are some arguments supporting the possibility that equity valuations may remain elevated or even richen further. There may have been a structural change that keeps real yields low and inflation moderate for at least another 5–10 years—perhaps a slowdown in equilibrium growth rate or a secular private sector deleveraging following decades of rising leverage. Or larger

saving pools and investors' better access to global capital markets at lower costs may have sustainably reduced the real returns investors require on asset class premia, and we'll never see a reversal. (A lower equity premium versus cash going forward would also help reconcile an academic puzzle of an "inexplicably" high excess equity return in historical data, the so-called equity premium puzzle.) We simply do not know.

There are also plausible arguments[28] that the CAPE is no longer an accurate measure of valuation relative to its own long history, and that adjusted measures give more neutral valuation signals in recent years.[29]

Whether or not valuations fall, lower real yields may justify a cautiously negative tactical view at the time of writing. Historical data suggest that a simple momentum signal may help to moderate that view in an industry of peer comparisons where early often equals wrong.

9 Concluding thoughts

In this paper we have examined how drifting valuations contribute to the initially puzzling gap between encouraging in-sample evidence and disappointing out-of-sample performance of value timing strategies. We have shown that the challenge of such long-term drift has been compounded by persistent negative exposure to a shorter-term time series momentum factor, which offers a practical and intuitive avenue for improving value-based tactical timing signals.

Some institutional investors instinctively prefer contrarian to momentum market timing. For an investor with a long horizon and correspondingly high tolerance of short-term losses, Warren Buffett's advice to be "fearful when others are greedy and greedy when others are fearful" rings true. The even snappier "buy cheap assets" is a persuasive and pleasingly concise maxim.

We too are value investors in many contexts, but the evidence challenges the idea that valuation signals *alone* can be used to time markets or inform asset allocation decisions. When others seem greedy, they may still get greedier for many years to come ("longer than you can remain solvent"). Even if the investor has the patience to stay the course, boards or capital providers, seeing persistent underperformance, may not. Keep in mind we're not talking about suffering through a few years in the wilderness, but decades on up to half a century!

Contrarians may be characterized as bravely or wisely standing up to herd-like market behavior. But while individual contrarian trades may indeed be uncomfortable to the point of heroism, the concept of "buying cheap" is so comforting and appealing—and *ex post* contrarian narratives are so misleadingly compelling—that it may be over-represented in tactical timing decisions to the point of sinfulness. For every peak or trough there will be investors who called it right, and subsequently attract publicity and praise. Many more call it wrong and fade into obscurity.

Conversely, momentum investing may feel too much like jumping on a bandwagon. Also, faster-moving momentum signals may seem unsuitable or impractical for institutions whose investment decisions are implemented infrequently and with a considerable lag and outlay of oversight resources. But our results suggest that even a small dose of momentum, whether applied explicitly or just by rebalancing less frequently or smoothing valuation signals, may improve market timing decisions.

In summary, attractive predictive correlations do not always translate to successful timing out of sample. This has occurred both because of long-drifts in average weight (as stocks become more or less cheap for very long periods) and because a pure value strategy effectively "shorts" the

successful momentum strategy. We have no cure for long drifts in average weights other than our view it's unlikely to be as bad for valuation timing in the next half century as in the last. In contrast the cure for shorting momentum is clear. Adding a momentum component may help investors to become more patient and successful long-term value investors.

Finally, we note that even a complementary blend of value and momentum timing signals produces only modest long-term outperformance. Thus: sin (only) a little.

Appendices

Strategy variants

Table A.1 shows hypothetical performance for the strategies (in bold) that we describe in the main article, and several other variants:

- "5-year Reversal" tilts are based on the past 5-year arithmetic return in excess of cash.
- "*EP* 60-yr window" is the value timing strategy discussed in the main article. The signal is based on data from 1881, with an expanding window until 1941 and then a rolling 60-year window.
- "*EP* in-sample" calibrates the signal based on the full 1900–2015 sample as shown in Chart 1, and therefore includes a hindsight bias.
- "*EP* Expanding" uses an expanding data window from 1881.
- "*EP* Extremes Only" ignores all tilts smaller than ±25% (variant 1) or smaller than ±40% (variant 2). The latter variant outperformed continuous contrarian signals during the more recent sample, but gives us very few

Table A.1 Performance of various timing signals in U.S. equities, 1900–2015.

	1900–2015				1958–2015			
	Excess return	Volatility	Sharpe ratio	Avg posn	Excess return	Volatility	Sharpe ratio	Avg posn
Buy-and-hold	6.6%	17.5%	0.38	100%	5.5%	14.9%	0.37	100%
5yr Reversal	6.1%	18.4%	0.33	95%	5.9%	17.4%	0.34	110%
***EP* 60yr window**	7.4%	20.0%	0.37	102%	5.4%	14.6%	0.37	89%
EP in-sample	7.8%	20.4%	0.38	105%	5.6%	15.4%	0.36	96%
EP expanding	7.5%	20.3%	0.37	105%	5.5%	15.3%	0.36	94%
EP extremes only 1	7.3%	20.2%	0.36	104%	5.8%	15.3%	0.38	95%
EP extremes only 2	7.4%	20.1%	0.37	107%	6.3%	15.6%	0.40	103%
EP - Nom. Bond Yield	7.8%	19.9%	0.39	98%	5.4%	12.7%	0.42	78%
EP - real bond yield	8.0%	20.7%	0.39	107%	5.7%	14.4%	0.40	88%
EP annual rebalance	7.7%	19.7%	0.39	102%	5.5%	14.3%	0.39	90%
EP 10yr lock-in	7.2%	18.9%	0.38	102%	5.5%	14.0%	0.39	90%
Momentum 12m	8.1%	17.1%	0.48	102%	5.9%	13.8%	0.43	97%
Momentum 3m	7.4%	17.3%	0.43	100%	6.1%	13.7%	0.44	97%
Value & momentum	7.8%	18.0%	0.43	102%	5.7%	13.8%	0.41	93%

Sources: Robert Shiller's Data Library, U.S. equity market returns from Global Financial Data, Ibbotson/Morningstar and Datastream. Hypothetical performance excess of cash, gross of transaction costs and fees. Arithmetic rates of returns and Sharpe ratios.

observations (overweight 1979–1984, underweight 2000, otherwise neutral).
- "*EP* – Nom. Bond Yield" signal is based on Shiller *EP* minus 10-year nominal bond yield, a version of the so-called "Fed model."
- "*EP* – Real Bond Yield" replaces the nominal yield with the yield in excess of forecast inflation, a more meaningful operation since *EP* is a measure of real return.[30] These relative valuation signals perform somewhat better than absolute *EP*, though they lose much of this advantage when combined with momentum (absolute Shiller *EP* is a better diversifier to momentum).
- "*EP* Annual Rebalance" rebalances annually at year-end instead of monthly—this is a crude way to lessen the fight against short-term momentum.
- "*EP* 10-Year Lock-in" uses a 10-year moving average of the signal, which is equivalent to investing 1/120th of assets based on the latest signal each month, locked in for 10 years (see also Chart 7). For example, this strategy benefits by locking in overweights during the 1980s and delaying the premature underweight in the 1990s.
- "Value & Momentum" is the combined strategy shown in Table 5, with equal weights to Shiller *EP* and 12-month momentum.

Bond market timing

Timing the stock market has been called by some an investing sin not only because it is hard, but also because it is a narrow and concentrated bet. Timing two assets is a broader and potentially more promising strategy, as well as providing independent supporting evidence.

We test the performance of simple value and momentum strategies applied to 10-year U.S. Treasuries, using as our value indicator the real bond yield, which is the nominal yield minus a survey-based forecast of long-term inflation.[31] Our momentum indicator is the 12-month excess return, as for equities, and we calculate the tilts using exactly the same process as before.

Results are shown in Table A.2 and Chart A.1. As with equities, a purely contrarian timing strategy fails to significantly outperform a passive buy-and-hold approach on a risk-adjusted basis. Using a combination of negatively correlated

Table A.2 Performance of buy-and-hold and simple timing strategies in U.S. Treasuries.

	1900–2015				1958–2015			
	Buy-and-hold	Value timing	Mom timing	VM timing	Buy-and-hold	Value timing	Mom timing	VM timing
Excess return	1.2%	1.4%	1.9%	1.7%	2.1%	2.6%	3.1%	2.8%
Volatility	5.9%	6.8%	6.1%	6.2%	7.9%	9.4%	8.2%	8.5%
Sharpe ratio	0.21	0.20	0.32	0.27	0.27	0.27	0.38	0.33
Max Drawdown	−21%	−29%	−13%	−18%	−21%	−29%	−13%	−18%
Max Relative DD		−29%	−8%	−7%		−29%	−8%	−7%
Avg position	100%	86%	103%	95%	100%	107%	104%	106%
Avg Posn Drag		−0.2%	0.0%	−0.1%		0.2%	0.1%	0.1%

Sources: U.S. Treasury returns from Global Financial Data, Ibbotson/Morningstar and Datastream, forecast inflation as in Ilmanen (2011). Hypothetical performance excess of cash, gross of transaction costs and fees. Arithmetic rates of returns and Sharpe ratios.

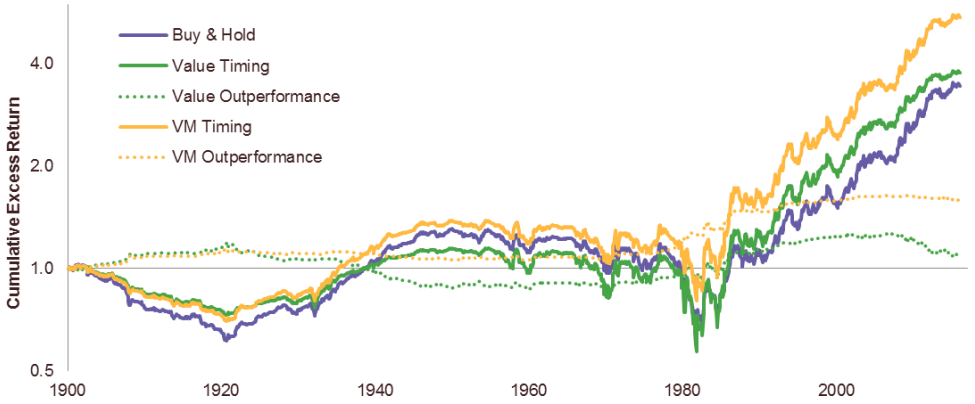

Chart A.1 Hypothetical cumulative return of buy-and-hold and simple timing strategies in U.S. Treasuries.

Sources: U.S. Treasury returns from Global Financial Data, Ibbotson/Morningstar and Datastream, forecast inflation as in Ilmanen (2011). Hypothetical performance excess of cash, gross of transaction costs and fees.

value and momentum signals earns higher returns, a higher Sharpe ratio and smaller drawdowns than either buy-and-hold or contrarian timing (though again momentum alone seems to be the outright victor).

We could continue to extend the idea of combining value and momentum tilts to other geographies and asset classes, and to cross-sectional or relative value strategies within each asset class, but that would be mission creep. Here we consider the practical challenges of timing a single market.

Notes

[1] We remember the late economist Paul Samuelson saying near the height of the technology bubble of 1999–2000 something along the lines of "market timing is an investing sin and for once I recommend that you sin a little." He meant, if he ever actually said it (we can't find a trace of it now on the internet), that things were so obviously wrong at that time that even a lifelong proselytizer of buy-and-hold would recommend – propitiously, as it turned out – some judicious selling.

[2] Since at least Campbell and Shiller (1988), and Fama and French (1988).

[3] Fisher and Statman (2006) compare the terminal wealth of stocks/cash switching strategies with 100% stocks, and find that in pure return terms the latter is hard to reliably beat. Estrada (2015) notes that buy-and-hold also has simplicity and lower costs in its favor.

[4] Another tactical advocate, Pfau (2011) suggests that a fixed 50/50 portfolio is a fairer benchmark for appraising strategies that allocate between stocks and cash, and that a more graduated signal is more realistic than full switching between the two assets. He also notes the influence of starting and ending valuations on performance during any given sample, and the importance of performance metrics beyond pure returns. After these adjustments, he concludes that valuation-based timing can be reliably preferable to buy-and-hold.

[5] The CAPE (cyclically-adjusted price-to-earnings) ratio uses average earnings per share over the past decade in the denominator to smooth cyclical variations in earnings. Both P and E are adjusted for inflation. Professor Robert Shiller popularized this idea and updates one version of the series regularly in his website. Our version is constructed slightly differently from that of Shiller – accounting for earnings publication lag, avoiding interpolation and the associated look-ahead bias, and using month-end rather than month-average prices. Results are similar for both.

Note that the CAPE uses smoothed earnings but current prices. Some studies have smoothed both earnings and prices, which effectively introduces a momentum component by implying more attractive valuations after periods of rising prices, compared to the standard CAPE. The deliberate use of smoothed signals to introduce momentum is discussed later. The related implications of using lagged prices in cross-sectional equity value factors are studied in Asness and Frazzini (2013).

[6] For similar evidence, see Chapter 8.6 in Ilmanen (2011): *Expected Returns* as well as Asness (2012): *An old friend, the stock market's Shiller PE*. Similar patterns are observed with total returns and real returns.

[7] In the early decades of the 1900s, the window is expanding using data since 1881. The choice of too short a window might move us away from the concept of fundamental valuation towards medium-term reversal, as our anchor becomes increasingly dynamic. For example, Bunn *et al.* (2014) use only a 20-year window to calculate their Relative CAPE Ratio, which can therefore be dominated by recent peaks and troughs. A 60-year sample seems long enough to represent a fundamental anchor, while allowing for some long-term trends in required risk premium or accounting practices.

[8] Goyal and Welch (2008) and Campbell and Thompson (2008) calculate an OOS *R*-squared that compares errors from OOS regression forecasts to those from an OOS mean. Dimson *et al.* (2013) regress subsequent returns on a series of out-of-sample forecasts, effectively comparing errors from a linear transformation of the OOS regression forecasts with those from the full-sample mean. We found the results, while generally weaker than in-sample measures, to be highly dependent on the OOS parameters (expanding or rolling window) and the choice of in-sample or OOS mean. By contrast, the backtest results that we present in the next section are similar for many reasonable specifications, as shown in the Appendix.

[9] While levering an equity portfolio may be challenging for many investors, now and (especially) in the past, our approach is equivalent to tilting an equity allocation above or below a benchmark: for example, if applied to a 60% allocation benchmark it would not require leverage. It has the advantage of allowing us to cleanly distinguish the passive equity premium and the timing view overlay, and to discuss any net under- or overweight over any given period.

[10] Note that we translate our signal to a notional tilt, with no risk targeting. Applying a risk forecast to the tilt or indeed the whole investment may provide improved time diversification, but is beyond the scope of this article.

[11] While Sharpe ratio is an imperfect measure for a strategy with changing risk through time, we show in Table 6 that alternative measures such as maximum drawdown lead to similar conclusions.

[12] The promising pattern of 10-year returns in Chart 2 persists even in this latter period (not shown).

[13] In other words, we don't know for sure that we are at a peak or trough until afterwards. Importantly, security selection or relative value strategies bypass most of this difficulty. A cheap stock can certainly get cheaper, and the attractiveness of a relative value opportunity must also be judged against the past, but we can at least confidently say that, by our chosen measure, one stock is currently cheaper than another.

[14] In fact, this result (and the timing performance in general) has been softened by another important source of hindsight bias: the choice of CAPE as a signal. Dividend yield was the most popular valuation signal at that time, later replaced due to the structural change of firms increasingly using buybacks instead of dividends. Dividend yield would have given an even more premature sell signal. Permanent structural changes present the worst outcomes for contrarian strategies: not only is the timing of the expected normalization difficult to judge—it may never happen.

[15] Another contributing factor is that without a hindsight bias the signal tends to linger near the extremes of its range, as can be seen in the graph of the signal in Chart 3. This exacerbates the tendency of time-varying risk to produce higher full-sample volatility and lower Sharpe ratios. See for example Kritzman (1999) and Hallerbach (2012).

[16] Following Bernstein (1997), who used a similar method for estimating expected returns. The periods may be overlapping.

[17] If there is a richening drift in the initial calibration period, the strategy will start underinvested and categorize the whole investment period as somewhat expensive, even though the CAPE is the same at the beginning and end.

[18] See Moskowitz *et al.* (2012) and Hurst *et al.* (2012). The latter paper notes that momentum exhibits attractive empirical tail-hedging-like behavior as well as positive returns. Note that trend-following strategies usually combine time series momentum strategies on many different assets, which is a much more diversified approach than the single-market momentum factor that we examine here.

[19] That is, looking back 60 years, the momentum factor is long (short) when the past 12-month excess return is above (below) the median, using the same scaling function as described for the value timing signal.

[20] Asness *et al.* (2013).

[21] We choose to omit estimated transaction costs from this analysis in order to focus on the fundamental challenges

of beating buy-and-hold, rather than implementation factors. The costs of applying a timing tilt would have been substantial in the past, but they are small today (an overlay of index futures could be used to express the tactical view).

[22] For very large investors, market impact considerations may point to giving a larger weight to value.

[23] The negative correlation gives a much lower tracking error for the combined tilt (3%) than for value or momentum alone (5–6%). For TE-unconstrained tactical investors, the in-sample optimal weight for the combined tilt (versus the passive portfolio) is higher than that shown in Table 5.

[24] Return correlations can differ from signal correlations if volatility also varies with the strategy signals.

[25] An example may be helpful. An investor seeking to benefit from return predictability over a 10-year horizon could in principle apply the timing signal only once a decade (in 1900, 1910, 1920, and so on), locking in their tactical tilt for the next decade. However, it is hardly realistic to expect them to ignore interim changes in market valuations, and for any empirical analysis the starting point is arbitrary. A more efficient and realistic approach involves investing a small fraction of wealth periodically (say, 1/120th every month) based on the latest timing signal, and holding a portfolio of 120 such overlapping 10-year locked-in positions. This strategy is equivalent to using a 10-year moving average of the timing signal to update the whole portfolio every month. The right-most point in Chart 7 depicts the performance of such a strategy; also see Table A.1, fourth row from the bottom.

[26] Yet another way to be patient, which may be unavoidable for some institutional investors, is to lag the signal. Lagging the value signal results in a similar improvement to that shown above, again by reducing the momentum headwind. A momentum signal lagged by one quarter remains a positive contributor; for longer lags its usefulness decays.

[27] For comparison, a corresponding strategy for U.S. Treasuries (see Appendix) has a larger underweight, driven by a very expensive valuation and near-neutral momentum.

[28] See Siegel (2016) and discussion in Philips and Ural (2016).

[29] One popular adjustment is to subtract the Treasury yield from the Shiller *EP* (or other *EP* measure), a version of the so-called "Fed model." Asness (2003) argued that the comparison is improper, but showed some short-term timing ability that we confirm in the Appendix. Low bond yields at the time of this writing make such a signal more bullish than the unadjusted Shiller *EP* we use as our base case.

[30] See Asness (2003).

[31] Before survey data are available, we use statistical estimates based on realized inflation as in Ilmanen (2011). Another candidate indicator for bond market timing would be the yield curve slope, or other measures of carry. Here we focus on value and momentum signals.

References

Asness, C. S. (2003). "Fight the Fed Model," *The Journal of Portfolio Management* **30**(1).

Asness, C. S. (2012). "An Old Friend: The Stock Market's Shiller P/E," *AQR White Paper*.

Asness, C. S. and Frazzini, A. (2013). "The Devil in HML's Details," *The Journal of Portfolio Management* **39**, 49–68.

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). "Value and Momentum Everywhere," *The Journal of Finance* **68**(3), 929–985.

Asness, C. S., Ilmanen, A., and Maloney, T. (2015). "Back in the Hunt," *Institutional Investor Magazine*, November 2015.

Bernstein, P. L. (1997). "What Rate of Return Can You Reasonably Expect. . . or What Can the Long Run Tell Us about the Short Run," *Financial Analysts Journal* **53**(2), 20–28.

Bunn, O. D., Staal, A., Zhuang, J., Lazanas, A., Ural, C., and Shiller, R. (2014). "Es-cape-ing from Overvalued Sectors: Sector Selection Based on the Cyclically Adjusted Price-Earnings (CAPE) Ratio," *The Journal of Portfolio Management* **41**(1), 16–33.

Campbell, J. Y. and Shiller, R. (1988). "Stock Prices, Earnings and Expected Dividends," *The Journal of Finance* **43**(3).

Campbell, J. Y. and Thompson, S. B. (2008). "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?," *The Review of Financial Studies* **21**(4), 1509–1531.

Dimson, Elroy, Marsh, P., and Staunton, M. (2013). "Mean Reversion," in *Credit Suisse Global Investment Returns Yearbook 2013*.

Estrada, J. (2015). "Multiples, Forecasting and Asset Allocation," Working Paper.

Fama, E. F. and French, K. R. (1988). "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics* **22**, 3–25.

Fisher, K. A. and Statman, M. (2006). "Market Timing in Regressions and Reality," *The Journal of Financial Research* **29**(3), 293–304.

Goyal, A. and Welch, I. (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *The Review of Financial Studies* **21**(4).

Hallerbach, W. G. (2012). "A Proof of the Optimality of Volatility Weighting over Time," *Journal of Investment Strategies* **1**(4).

Hurst, B., Ooi, Y. H., and Pederson, L. H. (2012). "A Century of Evidence on Trend-Following Investing," *AQR White Paper.*

Ilmanen, A. (2011). *Expected Returns*, Wiley.

Kritzman, M. P. (1999). *Economics and Portfolio Strategy Newsletter.*

Moskowitz, T., Ooi, Y. H., and Pedersen, L. (2012). "Time Series Momentum," *The Journal of Financial Economics* **104**(2), 228–250.

Pfau, W. D. (2011). "Revisiting the Fisher and Statman Study on Market Timing," Working Paper.

Philips, T. and Ural, C. (2016). "Uncloaking CAPE: A New Look at an Old Valuation Ratio," Working Paper.

Siegel, J. J. (2016). "The Shiller CAPE Ratio: A New Look," *Financial Analysts Journal* **72**(3), 1–10.

Keywords: Market timing; tactical timing; equity premium; Shiller CAPE; contrarian; momentum