Cliff's Perspective

—

The Long Run Is Lying to You[1]

March 2021

Everyone knows the value strategy has been a grave disappointment out-of-sample since, say, 1990 (thirty years!). Even Fama and French know it.[2] And everyone knows the value strategy has been quite bad for the last 10+ years with the 2018–2020 period essentially a crash. Well, that's all kind of true. But, as odd as this might sound, the realized average return on a strategy[3] is not necessarily the best estimate of its true long-term expected return. In fact, the right estimate of the true long-term expected return of the value strategy is considerably higher than many might think if they were to just look at simple past returns – especially right now.[4] Why?

Well, because one of the major things that buffets realized average returns is changes in valuations. To state the obvious, when strategies get more expensive, all-else-equal they do better while that richening is occurring – not necessarily afterwards when it's starting out expensive! Five years ago Asness (2016) followed Arnott, Beck,

[1] Because it's not nearly long enough to be immune to false readings from the valuation changes we'll describe below.

[2] They find that the value premium in the second half of the July 1963–June 2019 sample is statistically indistinguishable from zero, but because we have low power to estimate the mean return to low Sharpe ratio strategies, they also can't say it's statistically reliably lower in the second half. That may sound odd but in plain English they're saying "yeah, it's really low in the second half, but given its estimated efficacy it's still not a giant statistical shock." Oh, and Fama and French's results didn't even include the 2020 value investing fun…

[3] This is true whether in- or out-of-sample and even over seemingly long periods of time.

[4] We'll see later that even if one considers only the out-of-sample period commonly thought to be a failure for value, when estimated correctly, the long-term expected return to value you'd infer from this out-of-sample period is considerably higher than the disappointing realized average return.

Kalesnik, and West (2016) in examining strategy performance accounting for such valuation changes, and we extend that work here.[5,6,7]

Accounting for valuation changes when estimating long-term expected returns is not a new concept. Arnott and Bernstein (2002) do so for stock and bond market indices. Some other works that incorporate valuation changes[8] include Avdis and Wachter (2016), Claus and Thomas (2002), Cochrane (2006), Fama and French (2002), Grinold, Kroner and Siegel (2011) in this anthology, and in an exception that argues that valuation changes are in fact repeatable, Jetley, Neri, and Watson (2014). Still, while not brand new, I think this is a gravely underappreciated area of research. In particular, another underappreciated result is that when you estimate long-term expected returns accounting for valuation changes, you get a significant gain in the precision of your estimate (not just a different, less biased, estimate). How many of our well-known "facts" about expected returns, risk premia, and popular investment strategies are substantially influenced by prior valuation changes that are very dodgy to assume will repeat going forward? Not to ruin the suspense but the answer is a lot.

There has been far less work done on the implications for long-short strategies. Arnott, Harvey, Kalesnik, and Linnainmaa (2021) do a great job examining these strategies (see I can be nice!).[9] Here I examine some of the same things they do and also do some different things focusing on the regression methodology, which lends itself beautifully to dummy variable tests for various periods. In addition, this framework lets me stress a point that I think is grossly underappreciated, that valuation changes unnecessarily reduce the precision of our estimate of the true long-term market or factor expected return. Also, I focus not so much on value's 10+ year drawdown but on the out-of-sample since-1990 value factor returns (which many cite as proof that once a strategy is known it doesn't work anymore, or that the original value factor was datamined) and the really ugly value death of 2018–2020, while also including some different and novel market and strategy examples (e.g., intra-industry sales-to-enterprise value, EAFE vs. USA).[10]

Below I get to the value factor slowly, after looking at the same idea for the stock market (the most common place this issue has been discussed, see the links above), the bond market, and the more novel U.S. vs. international

---

[5] One of the major points of Asness (2016) is that Arnott et al. (2016) run some very interesting and useful regressions of strategy returns on valuation changes – but then promptly ignore these results in favor of other overstated measures that support their quite mistaken assertion that strategies like profitability, momentum, etc., have only worked historically because they got more expensive over time. In Arnott et al. (2016), despite actually presenting the regression results that tell us to always use less than 100% of the change in valuation in explaining contemporaneous factor returns, they instead use the full revaluation of value spreads over the whole period. So, if something got 100% more expensive over 100 years, forgetting all kinds of compounding stuff, they assume it would raise the annual average return by 1% per annum. But, they themselves show that the amount of revaluation that impacts returns is never equal to 100%. Rather, for value you get more like 90% of this revaluation passed through to the strategy (see the regression coefficients to come). This is what drove me batty in 2016 as, unlike value where it's only 10% mistaken (90% not 100%), when you make this conceptual error for momentum, BAB, profitability, etc., it is a giant mistake (way more than 10% mistaken, almost entirely wrong for momentum). Value is a low turnover strategy so ignoring this means using 100% of the valuation change to explain returns not 90%, a venial sin. But assuming 100% of valuation changes accrue to high-turnover strategies like momentum etc. is a cardinal sin. I can't do it all justice here, but please, if interested, read their 2016 paper and then mine. Happily, in Arnott, Harvey, Kalesnik, and Linnainmaa (2021) they do in fact use the proper regressions (though I had to find that in an internet appendix!), coming around to my way (just never acknowledging it which I, admittedly, begged for!).

[6] In fact it all got a little crazier (well, I got crazier, though I was also right – that particular combo of crazy/right is kind of my jam). Even after the 2016 exchange of papers they kept claiming, in subsequent work and conference presentations, that the long-term positive returns to factors such as momentum, profitability, and BAB were not real but were just artifacts of long-term richening (the exact kind of stuff we're going to study here except they got it wrong and willfully so). They did so even after I showed, actually kind of proved, that their own regressions (which, again, they ignored in favor of the clearly and obviously misleading methodology that assumes if valuations double over 50 years, a high turnover strategy like momentum benefits from all of that doubling even though that's clearly ridiculous as the portfolio is very different next year, let alone over longer horizons) told the opposite story. I called on them to either explain why I was wrong (inconceivable! – and yes that word means what I think it means) or retract their repeated false claims denigrating a generation of researchers' work on factors like momentum, profitability, and BAB. I am still waiting. But, as we'll soon discuss, their recent stuff is great (really! – that's not a sarcastic dig).

[7] Also in those early papers from 2016 and more in 2017, I argued against going big-time tactically extra long the value strategy, both because timing is inherently hard but also because, as of then, the strategy did not look particularly cheap versus history. Some thought it was super cheap and advocated loading up on it, but they were mostly either looking at just book-to-price (which was rather uniquely cheap at the time) or just realized losses, not whether those losses were justified. If you lose because of justifiable fundamental reasons you don't necessarily get cheaper. But, after arguing against jumping extra into value for a few years, in late 2019 I finally wrote that value spreads were at last so wide that it was time to sin a little (always our policy at true extremes rather than full abstention). And, as often happens with contrarian timing calls, certainly mine, I was a year too early, painfully so (COVID didn't help, but I wrote this before it really hit!). I'm still confident it will work out (though still not sinning a lot).

[8] Some of these studies come at it in another related way, incorporating contemporaneous fundamental change.

[9] While we have tracked this effect forever (since we were the first to define and explore value spreads!), I'm also happy for readers to consider this blog just an opinion (this is after all a blog and not the Journal of Finance!) that "hey they are right, value is awesome and only looks terrible because it's gotten so damn cheap, and that simple average returns are ***not*** the best forecast going forward."

[10] The EAFE one, in particular, is pretty cool!

stock comparison. I think all are pretty interesting and the intuition builds up before we tackle the long-short value factor (in several forms).

Let's Start With the S&P 500

Let's first look at returns on just the S&P 500 from 1950–2020. The following regression is annual excess (over cash) S&P 500 returns on the contemporaneous percent change in the Shiller CAPE.[11] That is, how much of each year's stock return is from valuation changes, not from, say, starting yield or realized earnings?

S&P 500 Return = 5.2% + 0.95 * change in valuation, R^2 = 93.2%

(8.5)   (35.9)

OK, 35.9 is a pretty high t-statistic, and 93.2% an impressive R^2. But we kind of knew that would happen going in, right? I mean, if someone told you in advance of every rolling year exactly how S&P 500 valuations (here just the CAPE but try it with your own favorite) would change, you'd have a pretty good trading strategy, right? Though it's still not a 100% R^2. I've already mentioned two reasons: starting yield and realized earnings – those drive a "wedge" in between valuation changes and realized returns. But there's a third. Even the S&P 500 has some turnover. So, when that turnover occurs, the return (LHS of regression) is the return on the actual S&P 500 each month, but the change in the Shiller CAPE (RHS of regression) is comparing a slightly different portfolio at the end of the period vs. the start.[12,13] The intercept is 5.2% with a t-statistic of 8.5. Of course, and I'm sorry in advance that I'll beat this to death below as a reminder, the strategy suggested by an 8.5 t-statistic is absolutely unimplementable. In the real world you have to bear the volatility induced by changing valuations.

The actual average excess return of the S&P 500 over that period is 6.5%[14] per annum versus the regression intercept of 5.2% above.[15] Why the difference? Well, because the CAPE has net gone up over the 71-year period (our version has moved from 10.9 to 34.5!). That was worth 130 bps of return per annum. That is, our estimate of the expected (if the CAPE were unchanged over the period) S&P 500 annual excess return is 130 basis points a year lower than its actual realized return (which, being real life, included the real life CAPE appreciation).[16] Now, if

---

[11] I use rolling overlapping years not months here, though either would work for the market return, as, for reasons explained later, annual returns make more sense than monthly when looking at the value factor return (ok, I'll explain here: the factor returns are rebalanced annually and this induces a big seasonal in value spreads that annual vs. monthly analysis avoids) and I want the regressions to be comparable. Of course, t-statistics are adjusted for the overlapping data. Also, the return on the left-hand-side (LHS) here is actually the continuously compounded annual return on the S&P 500 minus the continuously compounded annual return on T-Bills. The "change in valuation" on the right-hand-side (RHS) is actually the log of ending minus the log of starting CAPE (the continuously compounded return from just the CAPE change). It's important to use continuous compounding on the RHS as this preserves the property that if the CAPE ends at the same value as it started then the mean value on the RHS of the regression is zero (if it ended where it began but moved during the sample, its mean would be positive using simple compounding as, for example, +100% and -50% lead to an unchanged value but have a large positive arithmetic mean). If we used simple compounding, the regression intercept would not be precisely our estimate of returns to the stock market (or the bond market or the value factor below) in an unchanged valuation environment.

[12] You might ask why not compute the change in Shiller CAPE on the actual portfolio each year. That would indeed raise the R^2 a bit more (from the paltry 93.2%). But with the dynamic Shiller CAPE (the regular one) there is a strong, I'd say overwhelming, case that the expected percent change is long-term mean zero (that doesn't mean there aren't level shifts, but just that the long-term CAPE doesn't reliably drift up or down forever). That case is not as strong for a static portfolio. For instance, it could easily be the case that in months without turnover we expect some downward drift in the same-portfolio CAPE, but in months with turnover the opposite, as the S&P 500 perhaps replaces less expensive stocks with more expensive holdings. This difference would likely (my guess) be small for the S&P 500 but might be quite significant for higher turnover strategies like the value factor (the value factor is low turnover compared to many other dynamic factors but high turnover compared to the market index). And for our regressions to make sense the case that the valuation change is long-term mean zero must be strong. So, dynamic it must be.

[13] Ilmanen, Nielsen, and Chandra (2015) call all these "wedges" – the things that make changes in valuation less than a perfect explanation for realized return. In general they are much bigger for dynamic long-short strategies than for market index ones and bigger for higher versus lower turnover strategies. And turnover isn't the only major "wedge." Another example is beta hedging (if the long and short sides are different betas). Imagine you're long low beta stocks and short a considerably smaller amount of high beta stocks (to attempt to be zero beta). Imagine next that the market goes way up, low beta goes up more than you'd guess from its beta, and high beta goes up less than you'd guess from its beta; but, importantly, high beta still goes up more than low beta as that's what beta tends to mean in a sharply rising market means. Your long-short strategy made money but also got cheaper as the short portfolio went up more (a big wedge!).

[14] I'm taking the simple average of the LHS annual returns so they're directly comparable with the regression intercepts.

[15] Sadly that 6.5% only comes with a t-statistic of +3.5, a far cry from the +8.5 in the regression even though the regression intercept was lower (which, considered alone, would raise the t-statistic). Life is a lot harder when you don't have a time machine. More seriously, by adjusting for the contemporaneous valuation changes, we not only get a less biased, but also a less noisy, estimate of the long-run equity premium. Again, this doesn't make the actual investment any better, but it increases our confidence we're right about the steady state expected return of the investment (and that's very good!). So we've got that going for us, which is nice.

[16] Note that as I mentioned, I'm assuming that the expected future valuation change is zero. If you have a non-zero conditional expected valuation change, the intercept is not the full conditional expected return. But if the average conditional expectation of valuation change is zero

you were forming an estimate of mean returns in the future would you build a constantly increasing CAPE into your forecast going forward? I don't think so. So, while 6.5% a year is what you actually earned in the market, 5.2% a year is what you might long-term forecast going forward assuming neither mean reversion in CAPE (i.e., falling from its high ending value in our sample) or continued permanent expansion (you are allowed to assume either of these, of course). Either way the equity risk premium has been healthily positive, but it's 25% bigger if you, in my view wrongly, assume that the multiple expansion we've experienced is a permanent condition that should be built into future estimates. I think 5.2% is a better steady state forecast of the annual equity risk premium than the much more commonly used (actually nearly ubiquitously used) methodology of using simple realized returns not accounting for valuation changes that yields 6.5%.

But, again, note that even though our forecasted long-term expected excess return to the S&P 500 is 130 basis points lower accounting for valuation increases, it's also more than doubly as precise. I think the effect on our estimates of the mean, despite the noble efforts of work like Arnott and Bernstein (2002), is underappreciated. I think the effect of valuation changes on (unnecessarily) lowering the precision of our long-term estimate is in fact way more underappreciated, and is also quite important.

If we ran the same analysis again starting in 1950 but this time ending in March of 2000 (the peak of the tech bubble), it's even more extreme. The simple realized equity premium was 7.5% a year, but that was inflated by ending at a record CAPE of 45+. The estimate accounting for valuation changes (running the above regression but over this shorter period) is 4.9% a year (a tad lower than, but very close to, the full period results). It's a bigger difference here (2.5% between simple realized average return of 7.5% and the regression intercept of 4.9% vs. a 1.3% difference over the full period) because the period is somewhat shorter (these differences are magnified over shorter periods and would asymptote to zero if we had a big enough sample) and the CAPE was a fair amount more extreme at the top of the tech bubble than even today's lofty level. At times of extreme valuations, the standard method of estimating a return premium from historical (even over some pretty long periods) realized returns can be substantially off. Note the foreshadowing for the value factor as it is indeed at a ***very extreme level of cheapness*** today (I won't do hyperlinks this time – just see nearly anything I've written/whined about for the last year or two).

How About Fixed Income?

As another preamble to gain intuition, let's do bonds. We regress the annual excess over cash return on the U.S. government 10-year (levered or de-levered to a constant ex-ante duration of 7 years[17]) on the contemporaneous annual change in the 10-year yield using the same 1950–2020 sample period:

10-Year Bond Return = 2.0% - 6.3 * change in yield, R^2 = 95.4%

(9.4) (-39.4)

Now, unlike equities, you get almost the same estimate of annual bond excess return if you just only look at standard simple realized returns, rather than regress on the change in yields and then look at the intercept as in the above regression.[18] That is, simple average annual returns are 2.1% and the regression intercept above is 2.0%. Why so little difference here when we saw a significant one for the S&P 500? This is because over these 70 years, yields have taken an almost complete round trip with very little net change.[19] Therefore, the adjustment to estimated returns for their richening or cheapening is tiny (though the intercept is still measured much more precisely this way – still underappreciated!). Now, running the same numbers starting instead in July of 1984 (the beginning of the long secular bond bull market we're still in) through 2020, we find the average simple realized excess return of bonds over cash to be a whopping 4.4% a year (that is, 4.4% comes from the common exercise

over the full regression sample, I still think the intercept will be an estimate of the average conditional expected return. At this point that's a conjecture, not math, so clearly there is more to think about and do. I happen to think it's a very reasonable conjecture over the long term, but it still needs more work.

[17] We only have changing durations from 1980 onwards. Before that we use an estimated duration of 5.0. It doesn't really matter.

[18] Note, you get a pretty similar t-statistic on the intercept for bonds (the expected return if rates are unchanged) to that for the S&P 500 (the expected return if the Shiller CAPE is unchanged). T-statistics are proportional to Sharpe ratio, so the central contention of risk parity, that long-only stock and bond exposures have relatively similar Sharpe and thus should get relatively similar weight ***in terms of risk***, is supported here now accounting for valuation changes. Of course, this is just for USA stocks and bonds and not the main topic of this piece.

[19] Note, while I think these long-term regressions can make the reasonable assumption that the corresponding long-term average change in rates is zero, and are thus reasonable, at any point in time the shape of the yield curve combined with a number of different hypotheses (e.g., the expectations hypothesis) might make a richer study.

of just looking at realized bond returns vs. cash but ***not*** accounting for yield changes as in our regressions). If we run the above regression over this same 1984–present[20] period, we find:

10-Year Bond Return = 2.4% - 6.3 * change in yield, R^2 = 94.3%

(7.6)   (-32.9)

This 2.4% intercept in the July 1984–2020 regression is only a drop higher than the 2.0% in the 1950–2020 regression. That is because a whole lot of the giant (for bonds) 4.4% a year over cash from July 1984–2020 comes from the massive long-term fall in bond yields over this period.[21] Again, I don't think anyone should build that into their long-term estimate of bond expected return going forward.

OK, you may be thinking this is all pretty obvious. Of course, valuation changes explain a fair amount of short and even quite long-term returns (actually, that it matters over the long term wasn't that obvious ex ante). Still, many ignore this and use simple realized returns to forecast going forward expected returns. That is wrong if the expected change in valuation is zero going forward but was a non-trivial shock in-sample (as for stocks). All of this doesn't matter if valuation changes have been near zero over the period in question (like for bonds). But it can matter a lot if valuations have net changed a lot. In other words, in the above example one could use over 36 years (July 1984-present) of simple bond returns (not our regression framework but the normal exercise), think that's a very long sample, and confidently say the expected return of bonds in long-term steady state is double what it likely is in reality.

So, Now You're Dying to Know About the USA vs. EAFE, Right?

It's common knowledge that the stock market in the USA has crushed the rest of the world (I'll use EAFE here to proxy for non-U.S. developed market equities) since the peak of the Japan bubble (call that the end of 1989) or over the last, say, ten years, right? Well, yes. But here we are interested in how much of the outperformance is from the USA's fundamentals (or perhaps carry) beating the rest of the world versus the USA's valuations richening versus the rest of the world.

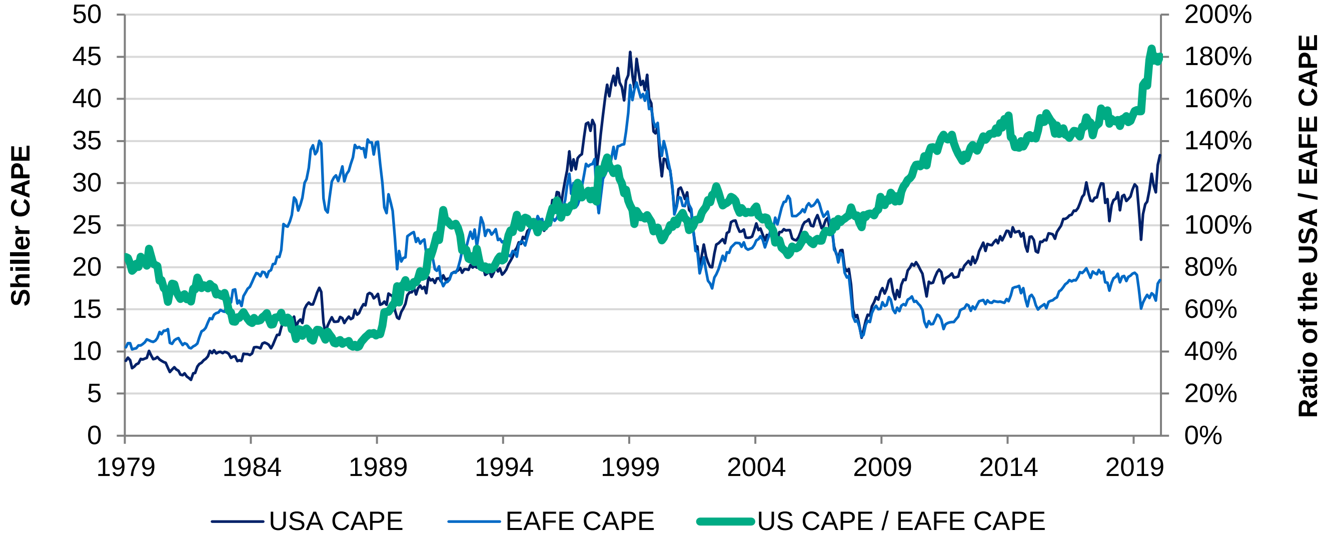
From 1980 to 2020 the USA has outperformed EAFE by 2.1% per annum.[22] A hefty number (though not statistically significant). Now, as you might guess, there have been some changes in absolute and relative valuation over this period. Below I graph the Shiller CAPE for MSCI USA (left axis in dark blue) and MSCI EAFE (left axis in light blue), and the ratio of the USA to EAFE CAPE (right axis in thicker green):

[20] Here and throughout "present" refers to through the end of 2020.

[21] While 2.4% is a better estimate of prospective long-run bond excess return than 4.4%, it may not be the best estimate at any given time if, for instance, the yield curve is flatter than average. But it should be a decent estimate of what investors were expecting, on average, during the past sample (while the average realized returns were boosted by presumably unexpected yield declines). It's also likely a reasonable estimate of the very long-term expected excess return on bonds (when the shape of today's yield curve or short-term forecasts of rate moves won't matter).

[22] We have EAFE starting in 1980 hence the shorter period. I switch from the S&P 500 to MSCI USA to match MSCI EAFE. The returns are approximations of currency hedged excess returns (continuously compounded). The approximation is they are local returns over local LIBOR (cap-weighted for EAFE) so they miss the cross-product term. Not a big deal at all.

USA vs. EAFE Shiller CAPE
December 31, 1979 – December 31, 2020.



Source: AQR, Bloomberg, Robert Shiller's Data Library, MSCI, Consensus Economics. Please see the Appendix for additional data information. For illustrative purposes only and not representative of a portfolio AQR currently manages. Hypothetical data has inherent limitations, some of which are disclosed in the Appendix.

We see a giant repricing with the USA getting relatively much more expensive on CAPE and ending effectively at a peak. You know what regression is coming right? I regress the difference in annual return (USA minus EAFE) on the contemporaneous difference in annual percentage change in CAPEs over 1980–2020:

USA minus EAFE = 0.4% + 0.94 * change in valuation, R^2 = 91.5%

(0.7) (19.6)

So, when we adjust for the change in relative valuation, the return differential shrinks from an economically (if not statistically) significant 2.1% per annum to an insignificant (both ways) forty basis points. The victory of the USA over EAFE for the last forty years is almost entirely coming from it getting relatively more expensive.

Now consider breaking the full sample into two subperiods. The first decade (call that "Japan Ascendant") 1980–1989, and then 1990–present (call that "USA Ascendant"). Over this first 1980–1989 decade, EAFE outperformed the USA by a whopping 6.9% per annum. So, let's run the above regression over only this Japan Ascendant period (yeah, only 10 years, have I mentioned this is a blog and not the Journal of Finance?):

USA minus EAFE = -0.8% + 0.96 * change in valuation, R^2 = 89.6%

(-0.6) (10.5)

Basically, the USA's loss goes away. It was all valuation change. Japan's valuation versus the world went mad and that explained it all.

Now, over the USA Ascendant period (1990–2020), the USA has outperformed EAFE by 4.4% per annum. This is almost as much as EAFE outperformed per annum in the 1980s, but it's even more economically significant as it occurred over a 3x longer period of time. Let's run our regression on valuation changes over this 1990–present period:

USA minus EAFE = 1.1% + 0.95 * change in valuation, R^2 = 92.3%

(1.5) (24.2)

It's still a 1.1% victory (my first obvious guess for why would be some real earnings outperformance in the USA) but a significantly smaller (1.1% vs. 4.4% or 75% smaller) victory then just looking at simple returns, and even though it's over thirty years it's not quite a statistically significant outperformance. The USA has won over the last thirty years mostly (75% in our estimate) on a revaluation from the late 1980s Japanese bubble and ending at the richest USA vs. world valuations we've seen. Building in an expectation that that will repeat, i.e., shunning international stocks because of the thirty-year drought everyone knows shows they stink, seems ill advised to say the least.

Basically, if you were estimating long-term expected returns going forward, would you want to build in CAPE appreciation going on forever at the rate in the graph above, or would you remove that effect and look at the intercept?[23] I think the answer is obviously the latter and hope by now you agree! This has real implications for asset allocators where the "removing the effects of massive changes in valuation" method argues for much more non-USA exposure than one would get just looking at simple realized returns and assuming they are a fair sample for the future (and this is without assuming any mean reversion from the historic high valuations of USA vs. EAFE). Basically, those who've changed their portfolio, selling their international stocks and putting it in the USA, because of the 30 years of USA outperformance, are kind of building in the assumption that relative CAPEs, which have gone up 2-3x for the USA vs. EAFE, do so again in perpetuity. Good luck.

Brief Review Before We Finally Look at the Value Factor

There are Two Separate Things Going On Here

Regressions of returns on contemporaneous valuation changes lead to both 1) a potential change in the level of the intercept versus average simple realized returns and 2) additional precision in measuring that intercept. These are related but still distinct effects. You can have one without the other (in particular, you can have an increase in precision from considering contemporaneous changes in valuation but no effect on the estimate if valuations end where they began). Frankly, I think 1) is underappreciated, but 2) is massively so.

For 1) to matter, the RHS of the regression, average valuation change, has to be non-trivially different from zero over the sample. If the average valuation change is zero, then no matter how much variance in return this valuation change contemporaneously explains, it won't change the intercept. The intercept will simply be restating the simple average realized return. But, if the valuation change is non-zero, particularly if the net change over the sample is extreme (a function of end points) then accounting for this change in our regressions can cause the regression intercepts to materially differ from the simple average realized return. I argue that when materially different, this new estimate is a better estimate of the long-term expected return of the factor, as effectively the realized return was under- or over-estimated because of extreme changes in valuation that should not be built into future expectations.[24]

Of course, contemporaneous valuation changes can explain a lot of realized return variance. Thus 2), we also get more precision in our estimate of the intercept. This doesn't mean we suddenly have a higher Sharpe ratio market portfolio or strategy. No, again, in real life we must live with this valuation-change-induced variance. But it can mean, and I think does mean, that even if the Sharpe is the same, we are more certain (more precise in our estimate) of the Sharpe's numerator (expected return) and thus the Sharpe itself. This is quite important as we aren't just concerned with finding the highest backtested Sharpes but, rather, finding Sharpes that are real and not the product of data mining or exceptionally biased sample periods. In English, being more sure you're right about the mean – for markets, market differences, and dynamic long-short strategies – is very good.[25]

Now, this technique is limited. It is most effective for very low turnover strategies. The stock and bond markets indices above clearly fit the bill. But for very high turnover strategies the "wedges" between valuation changes and realized returns are much bigger, and thus this technique is far less effective/important. Intuitively, if you have a super high turnover strategy, the ending versus starting valuation of its holdings can't matter very much as they are constantly different portfolios throughout the sample.[26]

Some famous factor strategies are high enough turnover to make this technique, perhaps not useless, but pretty weak. Momentum is one example (actually, here I'd really call it useless). But the value factor, while certainly

[23] By long-term expected return, I mean unconditional expected return. By talking about long-term steady-state expected returns, I am assuming away any effects of CAPE expansion or mean reversion on going-forward expected returns.

[24] If you have an explicit economic case for building them into future expectations, or arguing that they weren't expected to be mean zero over the long-term past, go for it. It's possible and I think this research is just beginning. But I think my assumption here is quite strong.

[25] This all has clear implications for capital market assumptions used by asset allocators. They can, quite possibly, be more sure of the true long-term expected return (more precision) and get a forecast of that long-term expected return free of bias (which we've shown can be material).

[26] Asness (2016) shows that R^2 values for similar regressions on higher turnover, bigger "wedge" factors (e.g., momentum, BAB, profitability) are indeed much lower. For instance, these regressions don't change the estimated long-term expected return to momentum or BAB or profitability much at all (an unresolved bone of contention between Asness (2016) and Arnott, Beck, Kalesnik, and West (2016)).

higher turnover than the stock or bond market indices, is a relatively low-turnover cross-sectional factor. Thus, there's great hope that this methodology applies and might add insight.

With all of that said, let's now (finally) turn to value.

The Value Factor

First HML

From 1950–present[27] the HML (devil[28]) factor realized an annual return of 1.9%[29] with volatility of 13.4% for a rather blah Sharpe ratio of 0.1.[30] As we've mentioned quite a few times, we do not think simple cap-weighted book-to-price is the best way to implement the value factor (so we think this is a low-ball estimate of what a value strategy can do). Nevertheless, even at this fairly weak standalone Sharpe, combining this value factor 50/50 with simple momentum (UMD) yields a Sharpe over 0.8 over this same period.[31]

Of course, while a 1.9% annual return (with a Sharpe of 0.1) over 1950–present is pretty weak, that's rolling in the dough compared to the 1990–present period often used as out-of-sample evidence that the value factor was data mined, is too crowded now, a scam perpetrated by rogue academics and their Wall Street progeny, etc. From 1990–present the HML (devil) premium was, well, zero – 0.0% per annum (gross of costs!) with a t-stat also, you know, zero. It's not hard to see how the statement that "the value factor has not worked since its discovery in the late 1980s" has many convinced.[32]

But we can run similar regressions to the above for the value factor. Here, we regress the annual value factor on the annual continuously compounded change in the "value spread" we've analyzed so often in other work (starting here where I think we invented the value spread – I may have already mentioned this above :)). In this simple specification, the value spread is the ratio of the book/price of cheap stocks over the book/price of expensive stocks and the annual change in the log of this ratio is what's used in the regression below.[33] From 1950–present:

Annual HML (devil) return = 3.0% - 0.91 * change in value spread, $R^2 = 78.8\%$

(4.0)   (-16.0)

So, as you can see in the figure below, because the value spread has gone up over this period (in fact ending at an approximate tie for the highest ever with the 1999–2000 tech bubble[34]) this has lowered realized simple returns (you lose when you get cheaper and lose a lot when you end much cheaper).

---

[27] Actually because the book value figures in HML Devil are updated at the end of June, these regressions are run over the non-overlapping June-June periods starting in 1950 (so the first is from end of June 1950 to end of June 1951) and the last is just six months (end of June 2020 to end of December 2020).

[28] Using our Devil version of HML and our data, which is very close but not precisely the same as Fama and French's. Devil uses the current market price with a lagged book value, while Fama and French lag both. While we always advocate the Devil version, it's much more important here as the lack of lagging in Devil is necessary to see the relationship between returns and valuation changes. The lags in the Fama-French methodology obscure this relationship.

[29] Again, for these stats I am taking the average of the continuously compounded annual returns I use in the corresponding regression. For HML this includes one six-month period along with the annual returns (the final one). I keep that here for proper comparability with the regression. Same applies to the volatility.

[30] Doing this analysis on annual continuously compounded returns yields a somewhat lower Sharpe than the more usual analysis examining arithmetic monthly returns.

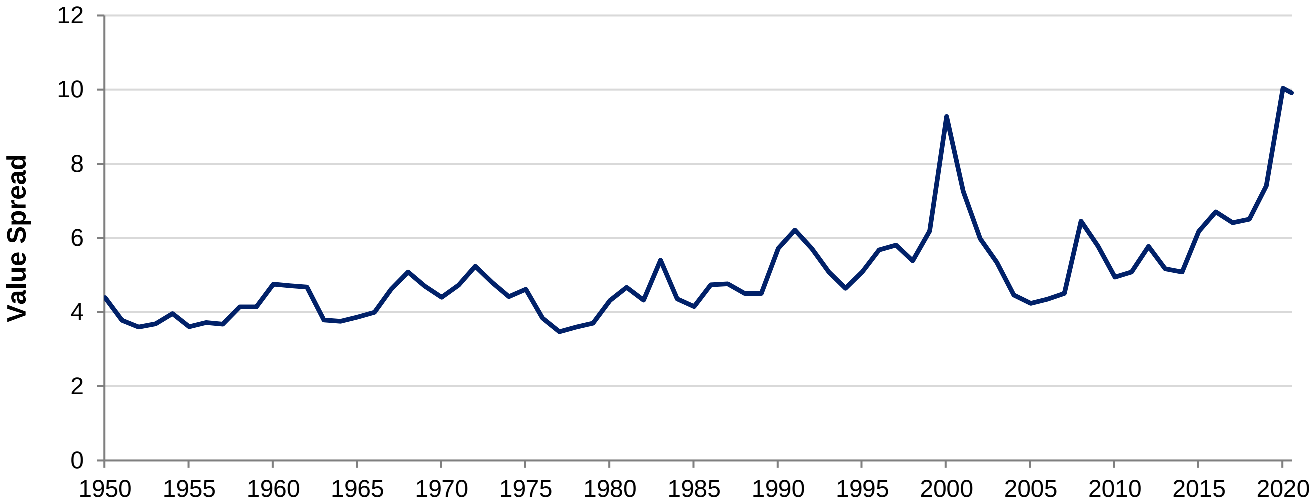
[31] Recall that another property of "Devil" implementation is a lower Sharpe standalone value factor (as it's more short the momentum factor, which hurts returns as momentum on average makes money) but a higher Sharpe combination with momentum. The standalone value factor here suffers from this univariately (the ugly 0.1 Sharpe) but benefits from it in combination with momentum. We won't discuss combinations with momentum any more in this note but, as usual, that's a big part of what makes the value strategy important.

[32] Combined with momentum it's still OK even 1990–present. Again, the Devil version generally presents a lower standalone Sharpe to value but a better Sharpe in combination with momentum.

[33] Note for the S&P 500 and USA-EAFE examples, I regressed returns on the contemporaneous percent change in the CAPE, so the coefficient in that regression was positive. For bonds I regressed on the contemporaneous additive change in the 10-year yield (all that's important for this exercise is for the RHS to be a measure of valuation and have a reasonable long-term expectation of mean zero) so because of how bond math works the coefficient is negative. Similarly for the value factor, I regressed on the percent change in the value spread, which we've always defined as the ratio of how expensive the expensive stocks are compared to cheap ones (so when it goes up the value strategy goes down and you get the negative coefficient in the text). Obviously, the signs are not important (if we negated any RHS measure we get the same result, just a different sign in the regression).

[34] In this case it looks a little wider than the 1999–2000 peak. Every way of constructing the value spread (e.g., What valuation measure? What universe? What weighting scheme? Do you allow industry bets or not?) of course yields a different value spread and value factor return

Value Spread, Book-to-Price
June 30, 1950 – December 31, 2020



Source: AQR, XpressFeed. Using the HML (devil) B/P value spread. Please see the Appendix for more details on the construction. For illustrative purposes only and not representative of a portfolio AQR currently manages. Hypothetical data has inherent limitations, some of which are disclosed in the Appendix.

This persistent cheapening over a long sample period is the opposite of the richening trends we saw for the equity and bond markets in the previous examples. Recall the average realized annual return to HML above was 1.9% from 1950–2020. But, the intercept, our estimate of the value premium had the value spread been unchanged (and not gone way up) is +3.0%. Already we're helping the value factor out. If you think the value spread cheapening (i.e., value getting cheaper and ending super cheap compared to history) is not the new permanent norm going forward (and if it is, you're forecasting perpetually widening, not just perpetually wide, value spreads), you'd estimate the value premium is over 50% more than what was actually realized (a 3.0% premium vs. 1.9%). Now, again, this by no means says value investors really made 3.0% out of sample! In real life you do in fact get buffeted by valuation changes and this can, and obviously does, impact even fairly long-term average returns (here it's seventy years!). But, what if our goal is not to report what happened, but discover our best estimate of the true expected return that the market had in mind when it set its valuations? If we believe, as I think quite obviously reasonable, that long term we expect no permanent drifts in valuation,[35] then 3.0% per annum is quite likely a better estimate of the going-forward expected value premium.[36]

Now let's look more at the dreaded 1990–present out-of-sample failure of the value factor. First, let's regress the full sample annual HML (devil) returns not on the contemporaneous value spread change but on a simple dummy variable that's zero through June of 1990 but positive afterwards through the end of the sample:

Annual HML (devil) return = 3.5% - 3.5% * post-1990-dummy, R^2 = 1.7%

(1.6)   (-1.1)

These results echo those of Fama and French (done somewhat differently) and our simple average returns reported above. We see that the post-1990 realized return is essentially zero (3.5% - 3.5%) but also that, statistically, we have little power to say it's actually ex ante lower post 1990 (t-statistic of only -1.1). This lack of power is testimony to the high variance involved.[37]

---

series. They are all correlated and telling a very similar story now, but which period is actually the widest changes from specification to specification (and I have noticed that most other specifications show a spike in the value spread late in the GFC that this one doesn't – oh well, again, correlated but not the same).

[35] Short-term you can still believe spreads will get even crazier, or, especially as the short term gets to the medium term, that spreads will mean revert from today's super high versus history levels.

[36] That's very long term. Again, at these levels of valuation we actually expect more than the normal steady state return over the next, say, 1–3 years and thus advocate sinning a little. But that's a forecast of mean reversion. No such forecast is necessary for any of the results in this essay to hold.

[37] Again, remember, we think other versions of value (intra-industry composites of multiple reasonable measures using weighting schemes other than value-weight) have higher Sharpes, and all the value factor Sharpes get higher when combined with momentum.

But now let's combine the above regressions (over 1950-2020) putting the annual value factor return on the left-hand side and, on the right-hand side, the annual value spread widening or tightening ***and*** a dummy for post-1990:

Annual HML (devil) return = 4.1% - 0.91 * change in value spread - 2.5% * post-1990-dummy, R^2 = 79.6%

(4.2)   (-16.1)   (-1.7)

So, how does our estimate for the ex ante value premium post- vs. pre-1990 change once we account for the portion of returns due to the value spread widening? Well, our out-of-sample post-1990 estimate is still lower than the earlier period. But rather than falling to 0.0% (3.5% minus 3.5%) it falls to 1.6% (4.1% minus 2.5%). Yes, it's still clearly lower than the pre-1990 period, but it's almost the same as the full period realized return of +1.9% we reported above (1.6% is still quite additive when you combine it with momentum).

Done another similar way, we run the regression on change in valuation over only the post-1990 period (instead of including a dummy we restrict the period):

Annual HML (devil) return = 1.6% - 0.93 * change in value spread, R^2 = 84.8%

(1.4)   (-12.7)

Looking at value over only the dreaded post-1990 period we find a return ***net of valuation changes*** again just about equal to the full 1950–present realized return. That is the return we would expect over this period ***if the value spread change is mean zero***. But going forward isn't that the estimate we want? Accounting for valuation changes, the estimated mean to even this relatively weak version of the systematic value factor is much better than typically reported. Post-1990 is now not a value failure but arguably a value success, just less than over the (valuation change adjusted) full period.[38] In other words, the post-1990 value failure is largely a result of changes in the value spread (value ending much cheaper vs. growth than it started), and if you think this spread doesn't permanently widen going forward, you aren't nearly as negative on the true expected return of value as those looking just at realized returns.

Now let's go crazy and look at the 2018–2020 terrible value crash. Regressing from 1950–present now not on a post-1990 dummy but a 2018–2020 dummy (call that post-2017) we get:[39]

Annual HML (devil) return = 2.4% - 11.9% * post-2017-dummy, R^2 = 3.2%

(1.5)   (-1.5)

Well, we kind of knew this, but wow. The value return is ***very*** negative over these three years (-11.9% per annum!) (though still only a -1.5 t-statistic over such a short period). But, now, you know what's coming, right? Let's add the spread change (again the regression is over 1950-2020):

Annual HML (devil) return = 3.0% - 0.91 * change in value spread + 0.4% * post-2017-dummy, R^2 = 78.8%

(3.9)   (-15.6)   (0.1)

Yes, this says that had spreads not widened (wish that were true!) you'd expect to have ***made*** 3.4% (3.0% plus 0.4%) over even this horrible period. The shocking part is the dummy is not negative but mildly positive. That is, if we didn't have this pesky massive cheapening of cheap versus expensive stocks, the value strategy (approximately) would've made (slightly and statistically insignificantly) ***more than usual*** over these last three years. This echoes results we've discussed elsewhere showing that the last three years' value debacle has been mainly (actually, in this case more than entirely) spread widening, not fundamental destruction.[40,41] Investors are

---

[38] Out-of-sample getting half the backtest is a win not a failure (some do confuse this!). If I were being glass half empty, I'd note that this is pre-costs and includes the Fama-French equal weighting of value within small caps. If I were being glass half full, I'd note that this is, again, a very simplistic measure of the value factor. But, either way, the out-of-sample evidence is far more consistent with the in-sample evidence after we account/control for valuation changes.

[39] Yeah, this is only 2.5 annual observations.

[40] The 2010–2017 bear market for HML was actually much more fundamentally driven than the last three years – an occurrence in which other forms of value, like those not taking an industry bet, fared better, and in which other factors, like profitability, various forms of fundamental momentum, etc., can help much more than when markets can be characterized by non-fundamental return chasing. Ironically for a manager betting not just on value but also on quality, momentum, low beta, etc., we think it's better for value to lose on the fundamentals (when the other factors besides value can help) than to lose in a mania (when other factors designed to pick up real value traps can't really help).

[41] Another potential problem is not whether the price moves were justified by the fundamentals or not (which is not a problem for this method,

simply paying much more for the same fundamentals (for expensive vs. cheap stocks) as opposed to fundamentals radically improving for the expensive relative to the cheap than they were at the end of 2017.

A Robustness Check on HML – Intra-Industry Sales-to-Enterprise Value

If you fully believe me after the HML[42] results above, you can safely skip this section. But if you don't trust anything with book-to-price involved (those meddling intangibles!), this should help. I do similar tests but with a very differently constructed value factor. I look at the sales-to-enterprise value (S/EV) factor and follow the methodology explained in part 3 of an earlier blog.[43] Here we're using S/EV but doing so industry-neutral, only among the top 1000 stocks, not cap-weighting, and making an adjustment to try and keep the long-short strategy ex ante beta neutral. We only have this data 1968–present[44] so instead of 1990–present being our out-of-sample period, I divide the data in half making July of 1994 to the present the new out-of-sample period.

The simple return of this S/EV factor over the full period is 2.9% per annum with a t-statistic of 1.9.[45] Now, of course, regressing on the contemporaneous annual change in the S/EV value-spread[46] is the next step:[47]

Annual S/EV return = 3.6% - 0.98 * change in value spread, R^2 = 52.1%

(3.4)   (-7.4)

First, notice the t-statistic on the contemporaneous change in spread and the R^2 are lower than for HML above. Many of the differences here vs. HML contribute to this amelioration. More important than using S/EV versus B/P, using equal weight long and short portfolios made out of the top 1000 stocks done industry-neutral[48] and beta-adjusted[49] all increase the "wedges." Still, the overall effect is very similar. The intercept is somewhat larger than the raw return (3.6% vs. 2.9%), indicating rising valuations for expensive versus cheap over this period that bias the simple returns down, but it is a bias that is fixed by looking at the intercept instead of simple returns. The precision goes up substantially, close to doubling (a t-statistic of 3.4 vs. 1.9). The estimate of the S/EV factor return is better adjusting for valuation changes and we're far more certain it is positive.

Now let's regress on the second half (July 1994–present) dummy to look at out-of-sample value performance (we certainly knew about sales-to-EV in 1994 so this is a fair out-of-sample test):

Annual S/EV return = 4.0% - 2.3% * post-1993-dummy, R^2 = 1.1%

(1.8)   (-0.7)

In the second half, S/EV realized +1.7% (4.0% - 2.3%). This out-of-sample period result isn't as low as we saw in HML (+1.7% vs. zero post-1990 for HML) and again there's not much explanatory power in a -0.7 t-statistic, but still the realized return was certainly lower than in the first half return of 4.0%.

But might spread widening be to blame? You guessed it, let's add in the valuation change to find out.

though it can affect our forecast going forward if one is so inclined), but if the indicator itself becomes very biased. The common example now is book value, which we use here, understating some firms' valuation (e.g., firms with a lot of "intangibles"). That is, yes, the returns are all driven by people paying more (less) for the glamor (doghouse) firms, but that is justified as the indicator is broken. In this case, even the returns caused by the value spreads widening (if due to book becoming biased) should likely still accrue against the long-term expected return. But, as we've shown in other work and below, this analysis is quite robust to using other valuation measures that are likely far less impacted than book value by such changes, and value spreads being super-wide today is not driven by the handful of potentially monopolistic companies that perhaps (perhaps!) might justify a higher P/B multiple. Also, value spreads are not being driven by the tech industry, the most expensive stocks, etc. All of this points to a very robust result that is not particularly beholden to book-to-price and its (potential) problems.

[42] The book-to-price factor, constructed as Fama and French do, save with contemporaneously measured prices.

[43] Refer to the section called "As Realistic as We're Going to Get Here."

[44] A somewhat shorter period versus 1950-present.

[45] Of course, adding in some momentum would make this t-statistic much stronger.

[46] A measure of how much cheaper or more expensive the new S/EV portfolio at the end of a year ended up versus at the start.

[47] By the way, there's no reason to have the RHS just be the change in one type of valuation (here S/EV). I've tried a few versions of regressing on multiple measures of the change in valuation (e.g., the continuously compounded change in the S/EV and B/P value spreads) and generally found even stronger results. Something for the future.

[48] We take the z-score of S/EV within industry and then sort on the entire cross-section.

[49] We divide the long and short sides of the portfolio by their respective ex-ante betas.

Annual S/EV return = 4.0% - 0.98 * change in value spread - 0.7% * post-1993-dummy, $R^2 = 52.2\%$[50]

(2.6) (-7.3) (-0.3)

Accounting for spread change, this second half sample fell only to 3.3% (4.0% - 0.7%) rather than 1.7%. This isn't too far off the first half realized return of 4.0% and is actually higher than the full sample realized return of 2.9%. This certainly doesn't seem like a value failure! Well, it was a failure for value investors as real-life counts! But, in a deeper way it is not evidence we should think less of value now than thirty years ago.[51]

Now, let's run the post-2017 dummy (this time there will be blood):

Annual S/EV return = 3.6% - 13.1% * post-2017-dummy, $R^2 = 7.4\%$

(2.3) (-2.0)

Over the 2018–2020 value destruction, our S/EV factor got clocked even worse than HML.[52,53] Of course, you know the drill now, let's include the spread change ***and*** the post-2017 dummy (1968-2020):

Annual S/EV return = 3.9% - 0.95 * change in value spread - 4.2% * post-2017-dummy, $R^2 = 52.8\%$

(3.5) (-6.9) (-0.9)

OK, over the 2018–2020 period of total value destruction S/EV did lose even after considering how much cheaper it got, unlike the simple book-to-price factor. But a -4.2% event is a far cry from a -13.1% one.[54]

Once again, much, even most of what we think of as strategy return is actually valuation change. Those matter as they affect the real-life strategy.[55] I promise I don't forget that! This value strategy did lose by -13.1% per year versus its pre-2018 average and not the -4.2% found after removing the effect of valuation changes. But, valuation changes shouldn't matter to estimating the true expected return of the strategy.[56] Again, net of valuation changes, the value premium (like the stock market, bond market, and USA vs. EAFE average returns) is far more stable and far more precisely estimated than the usual simple analysis would suggest.[57]

What if the Change in Valuation is Justified?

Does it change anything if the movements in relative valuation, which explain most of the variance in period-to-period returns (and in some cases radically alter the conclusions we make when looking at simple returns), are justified? That is, if the asset that got more expensive got more expensive for good rational reasons versus blowing up a big bubble, does it matter which of those happened?

It can clearly matter to your forecast of the conditional opportunity. If something got cheap because it deserved to get cheap, that could mean no extra expected return going forward is warranted (as opposed to getting cheap for

---

[50] And like with HML, we also ran the regression over just the post-1993 period, restricting the period rather than using a dummy (really almost the same thing as a dummy lets the intercept vary but the regression over a shorter period also lets the coefficient on the change in value spread vary). The results are essentially analogous: Annual S/EV return = 3.4% - 1.0 * change in value spread. The t-stats were 1.8 and -5.4, respectively, and $R^2 = 53.6\%$.

[51] As usual, unless you think that spread widening is a permanent condition and people are now fully rational.

[52] I've mentioned this in a few other places and in countless presentations but ironically B/P gets all the heat (it's so naïve, and the intangibles!!!) but, while it's been worse long term and during the value drawdown prior to 2018, over the true value crash (2018–2020) it's actually been one of the better (that is, less horrific) value factors. Go figure.

[53] If this were a published paper, and not a blog, I'd decompose which of the many changes from cap-weighted B/P to industry-neutral-equal-weight-done-only-in-large-caps-with-a-beta-adjustment S/EV is driving this – but it isn't, so I'm not.

[54] With less than 1/2 the statistical significance, a -0.9 std event can very easily be bad luck.

[55] Again, value managers really have suffered, they don't get to eat the intercepts.

[56] And you can't reject that S/EV isn't the same mean in the second half of the sample or even over the 2018–2020 value crash net of the effect of valuation change.

[57] Lest you think valuation can change all the simple results we hold so dear, I tried it for the superiority of value investing in Japan versus the USA, and it didn't explain anything at all. Value investing just kicks *** in Japan (seriously, we still wouldn't assume that this will always be true going forward, but it is what has happened and it's not because value got super relatively expensive in Japan). I've investigated this issue before, but starting from the highly (very highly) related phenomenon of momentum's failure in Japan. My conclusion was a bad draw for momentum / good draw for value in Japan. Boring and unsatisfying I know. By the way, I also tried to resuscitate the pure small cap premium (not the one net of quality that actually exists) checking if its disappearance out of sample was mostly because of small caps cheapening. But, alas, that wasn't to be (though I took one stab at it, I think it's worth looking at more).

no reason when perhaps you would expect higher expected returns going forward to your factor). But, I don't think it matters (though admittedly this is my first go at thinking this through and it is a bit hard to get your head around!) for our estimation of the true long-term expected return of the asset or strategy. Imagine expensive stocks are suddenly worth more because they should, for the same fundamentals, be worth more than in the past. For instance, assume there was a real and rational shock to our forward-looking long-term[58] earnings growth expectations. Clearly, in that case we'd see relative valuations move against the cheap and for the expensive (as we've seen). It would also mean today's super high value spread does **not** imply conditionally higher than normal value returns going forward. That part I can't save (though I don't believe the premise that it is all reactional).[59] But it would not alter our main conclusion – that the poor last ten or three years (or even post 1990) returns to value do ***not*** mean the long-term expectation for value is now much lower.

From this perspective, an irrationally high ending value spread is no different from a rationally high one. It's something that has happened, that has driven returns, but unless we forecast that spread widening will go on in perpetuity (not spreads staying at current high level but going ever wider), the intercepts of our regressions are still the better estimate of the true long-term expected return than simple realized returns. And, as we've shown over and over, those intercepts are much more stable than simple in-sample averages. The only way our long-term forecast changes is if we now believe the difference between expensive and cheap firms not only jumped radically, but in addition, individual stock prices are no longer influenced by the investor biases which have driven the very existence of the value premium long-term (i.e., the value premium has gone away).[60] That the shock to expensive versus cheap stocks is really justified this time is a bet we'd take the other side of (we do think value's conditional expected return today is higher than normal), but not bet the farm on (that's why we only sin a little!). That the level of prices today are set far more rationally than ever before is certainly possible, but is much closer to something we'd wager our arable land against.

So What Does All This Mean?

Basically, accounting for valuation changes makes us more sure that the stock market, the bond market, and the value factor's expected return is a healthy positive (even if we still have to live with the real-life volatility induced by valuation changes). Similarly, it makes us more sure that international stocks' mean return doesn't truly differ from that of USA stocks – at least not close to the degree found from simple returns. All of these have important implications for long-term asset allocation.

In particular, this analysis rescues value from the dustbin of history that many want to throw it in based on its 1990–2020 out-of-sample, or its even more recent, performance. This out-of-sample loss has spawned countless convenient, smart-sounding[61] stories about "data mining" and "once a strategy is known it stops working" and "oh, clutch my pearls, the intangibles!" that are the kind of narratives always shouted from roof tops ***after*** a large loss. Frankly, lightweight wannabes try to make their bones telling these stories ***after*** extreme movements. They do it every time. Only the specific players and rationalizations differ. The stock characters live on eternally.

Value is a much better strategy than many would, and do, think from just looking at realized returns that end at a time of ***extremely*** cheap value vs. growth prices. This should increase both our long-term confidence in the value factor and our staying power during its rough, sometimes very rough, periods (and, for that matter, it should do the same for the stock and bond markets as well as allocations outside USA stocks, when their realized returns are at big extremes in either direction – either increase our staying power after bad times driven by valuations or make us refrain from wildly overweighting after good times driven by the same). Right now, that means not ***over***estimating bond and stock market expected returns due to ending high valuations, not ***over***estimating USA versus international equity returns because of the same, but not ***under***estimating the value factor return because expensive stocks are now priced so high versus cheap ones. It would not be going too far to say that avoiding these misestimations, in both directions, and the portfolio misallocations they bring is a large part of the job of a

[58] Long-term is the rub here. Investors are notoriously poor at forecasting such changes out much more than a year.

[59] Of course, this is not our forecast. Our forecast is that the behavioral forces that lead people to pay too much for the stocks they love versus those they hate are the same as usual. And, while we don't trade on anecdotes, does today feel like a coldly rational market?

[60] Again, using only the behavioral story for value, but as usual a corresponding risk story works too.

[61] Stories ex post explaining why something bad has happened and obviously will now continue forever always sound smart as they are ex post (but often aren't – sometimes they are, so you need to keep an open mind, just not often). In many of my other blogs I stress the need to keep an open mind and explore all reasonable stories for why you might be wrong going forward. That is not the same as crafting ex post weak stories after a bad period because you know the world is dying to hear them then, and this is your chance for fame.

long-term investor. If we (the royal "we" that certainly includes us, but also you readers!) aren't doing this, we're simply not doing our jobs.

Most generally, I think studies like these that account for valuation changes should have a considerably larger role in our field, where most of the work is on simple (not valuation-change-adjusted) realized return (i.e., examining a factor's simple average return and simple average return net of its risk exposures). I could see this being a ubiquitous part of the researcher's toolkit (when would you not want to know if your results were biased in either direction by presumably non-repeating revaluation over your sample period?). This will matter most for strategies with low turnover (and relatively mild other "wedges") and big valuation changes over the period examined. But, as we see here when studying the value factor, the effect of adjusting for valuation changes can be quite dramatic even over a thirty-year sample period, in both our estimate of the true mean and (especially underappreciated) in the precision of this estimate.

Finally, an important takeaway is just how long a strategy (including stalwarts like the stock and bond markets) can deliver extra high (if valuations soar at the end) or disappointingly low (if valuations crater at the end) returns and ***not have it change*** our view of their long term expected returns. That makes all these strategies[62] hard to live with in real life. You can be wrong (or luckily right) for really long periods. But, such is the reality of investing, and of the business we've chosen (not that there are better choices for long-term investors!).

This essay can't change these facts. The difficulties of sticking with long suffering investments, even if they are just as good as they ever were, are real. As are the difficulties of avoiding falling in love with a bull market. Just because you understand value had a subpar thirty years or a disastrous last three years, mostly due to valuation changes (that you don't expect going forward), doesn't mean you didn't live through it. It also doesn't mean that people aren't clamoring for your head! But, perhaps, this analysis can make overcoming these difficulties and not abandoning good strategies[63] when they are very cheap or overweighting strategies when they are very expensive just a little bit easier.

[62] Again, this includes not just a dynamic strategy like the value factor but the stock and bond markets too (try to remember how you felt, and how many tried to explain to you that they were now bad ideas going forward, when they had terrible trailing results!).

[63] Or even abandoning rock-solid concepts like diversification.

Appendix

Data Descriptions and Sources

Equity (S&P 500) Regressions

Returns: Source: AQR, Bloomberg, Ibbotson, Datastream. Equity returns are the S&P 500 index. The S&P 500 Index, or the Standard & Poor's 500 Index, is a market-capitalization-weighted index of the 500 largest publicly traded companies in the U.S. Cash is 3-month T-Bills.

Shiller CAPE: Source: AQR, Bloomberg, Robert Shiller's Data Library, MSCI, Consensus Economics. We adjust Shiller data to reduce look-ahead bias and use month end prices. Universe roughly corresponds to S&P 500 TR for the U.S. Figures are subject to change.

Fixed Income (U.S. 10-year Government Bond) Regressions

Returns: Source: AQR, Datastream. Returns are 10-year U.S. government bond returns based on the J.P. Morgan US Government Bond Index (which consists of eligible fixed-rate, USD-denominated treasury bonds issued by the United States Government) and scaling to duration 7. Prior to 1980, duration 5 is assumed. Cash is 3-month T-Bills.

Yield: Source: AQR, Datastream. The yield is the U.S. government 10-year yield.

USA vs. EAFE Regressions

Returns: Source: AQR, Bloomberg. USA equity returns are the MSCI USA index. The MSCI USA Index is designed to measure the performance of the large- and mid-cap segments of the US market; it has 602 constituents covering approximately 84% of the free float-adjusted market capitalization in the US. EAFE returns are based on the MSCI EAFE index. The MSCI EAFE Index is an equity index which captures large- and mid-cap representation across 21 Developed Markets countries, excluding the US and Canada; it has 874 constituents covering approximately 85% of the free float-adjusted market capitalization in each country. Local LIBOR is based on monthly 3-month LIBOR Rates. For EAFE this is constructed using market cap weights of the MSCI EAFE constituents (dollar denominated market caps for single country MSCI indices) to weight the various country level LIBOR rates.

Shiller CAPE: Source: AQR, Bloomberg, Robert Shiller's Data Library, MSCI, Consensus Economics. We adjust Shiller data to reduce look-ahead bias and use month end prices. USA CAPE is based on the MSCI USA universe. Figures are subject to change. The EAFE CAPE is created by taking country-level data and weighting it according to the MSCI EAFE weights. Note that these MSCI universes are not tradable universes and also do not embed growth assumptions given that we do not have composite level growth assumptions at this time. Figures are subject to change.

Value: HML (devil) and S/EV

Data Sources: AQR, XpressFeed. Pricing and accounting data are from the union of the CRSP tape and the Compustat/XpressFeed Global database. The universe is all available common stocks in the merged CRSP/XpressFeed data.

Returns: The "HML (devil)" value strategy is the academic HML approach, using a book-to-price factor built over a U.S. all-cap universe that combines the NYSE, AMEX and NASDAQ. It is similar to the Fama-French HML factor, except that up-to-date prices are used. The S/EV factor is constructed using a universe of the top 1000 stocks in the U.S. We take a z-score of the S/EV within the industry and then sort the stocks in the entire cross-section. We go long the top 30% of stocks and short the bottom 30% using NYSE breakpoints. The long and short sides are equal-weighted portfolios. The industry classification is based on Fama-French 48 industries before 1986 and MSCI GICS groups after 1986.

Value spread: spreads are formed using the same factor as the underlying portfolio. So HML (devil) spread is the B/P spread of the B/P portfolio – so the B/P of the cheap stocks over the B/P of expensive stocks where cheap and expensive are defined by the B/P portfolio described above – and the S/EV spread is the S/EV spread of the S/EV portfolio.

Disclosures

The views and opinions expressed herein are those of the author and do not necessarily reflect the views of AQR Capital Management, LLC, its affiliates or its employees.

Past performance is no guarantee of future results.

Diversification does not eliminate the risk of experiencing investment loss.

This document has been provided to you solely for information purposes and does 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.

There can be no assurance that an investment strategy will be successful. Historic market trends are not reliable indicators of actual future market behavior or future performance of any particular investment which may differ materially and should not be relied upon as such. This material should not be viewed as a current or past recommendation or a solicitation of an offer to buy or sell any securities or to adopt any investment strategy.

Broad-based securities indices are unmanaged and are not subject to fees and expenses typically associated with managed accounts or investment funds. Investments cannot be made directly in an index.

HYPOTHETICAL PERFORMANCE RESULTS HAVE MANY INHERENT LIMITATIONS, SOME OF WHICH, BUT NOT ALL, ARE DESCRIBED HEREIN. NO REPRESENTATION IS BEING MADE THAT ANY FUND OR ACCOUNT WILL OR IS LIKELY TO ACHIEVE PROFITS OR LOSSES SIMILAR TO THOSE SHOWN HEREIN. IN FACT, THERE ARE FREQUENTLY SHARP DIFFERENCES BETWEEN HYPOTHETICAL PERFORMANCE RESULTS AND THE ACTUAL RESULTS SUBSEQUENTLY REALIZED BY ANY PARTICULAR TRADING PROGRAM. ONE OF THE LIMITATIONS OF HYPOTHETICAL PERFORMANCE RESULTS IS THAT THEY ARE GENERALLY PREPARED WITH THE BENEFIT OF HINDSIGHT. IN ADDITION, HYPOTHETICAL TRADING DOES NOT INVOLVE FINANCIAL RISK, AND NO HYPOTHETICAL TRADING RECORD CAN COMPLETELY ACCOUNT FOR THE IMPACT OF FINANCIAL RISK IN ACTUAL TRADING. FOR EXAMPLE, THE ABILITY TO WITHSTAND LOSSES OR TO ADHERE TO A PARTICULAR TRADING PROGRAM IN SPITE OF TRADING LOSSES ARE MATERIAL POINTS THAT CAN ADVERSELY AFFECT ACTUAL TRADING RESULTS. THERE ARE NUMEROUS OTHER FACTORS RELATED TO THE MARKETS IN GENERAL OR TO THE IMPLEMENTATION OF ANY SPECIFIC TRADING PROGRAM WHICH CANNOT BE FULLY ACCOUNTED FOR IN THE PREPARATION OF HYPOTHETICAL PERFORMANCE RESULTS, ALL OF WHICH CAN ADVERSELY AFFECT ACTUAL TRADING RESULTS.

“Expected” or “Target” returns or characteristics refer to expectations based on the application of mathematical principles to portfolio attributes and/or historical data, and do not represent a guarantee. These statements are based on certain assumptions and analyses made by AQR in light of its experience and perception of historical trends, current conditions, expected future developments and other factors it believes are appropriate in the circumstances, many of which are detailed herein. Changes in the assumptions may have a material impact on the information presented.

AQR Capital Management, LLC, (“AQR”) provide links to third-party websites only as a convenience, and the inclusion of such links does not imply any endorsement, approval, investigation, verification or monitoring by us of any content or information contained within or accessible from the linked sites. If you choose to visit the linked sites, you do so at your own risk, and you will be subject to such sites' terms of use and privacy policies, over which AQR.com has no control. In no event will AQR be responsible for any information or content within the linked sites or your use of the linked sites. Information contained on third party websites that AQR Capital Management, LLC, (“AQR”) may link to are not reviewed in their entirety for accuracy and AQR assumes no liability for the information contained on these websites.

Information contained on third party websites that AQR Capital Management, LLC, (“AQR”) may link to are not reviewed in their entirety for accuracy and AQR assumes no liability for the information contained on these websites.

This document is not research and should not be treated as research. This document does not represent valuation judgments 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. This document has been prepared solely for informational purposes. The information contained herein is only as current as of the date indicated, and may be superseded by subsequent market events or for other reasons. Nothing contained herein constitutes investment, legal, tax or other advice nor is it to be relied on in making an investment or other decision.