Russell Research

By: Don Ezra, Co-Chairman, Global Consulting

Insights for evaluating active management

The purpose of active management

In principle, an investor selects an asset allocation policy that reflects a trade-off between two extreme goals: to seek as high a return as possible, and to take as little risk as possible. The trade-off compromises both goals, to an extent that reflects the investor's risk tolerance. Once that choice is made, the investor can then implement the selected policy passively, by buying portfolios that match market indices in each asset class (or the nearest equivalent to a market index, for an asset class without an investable index). This is typically the most inexpensive way to secure market exposure to each asset class.

But most investors choose active rather than passive management. They hire one or more managers who take positions (which I'll call "bets") against their passive benchmarks. Since this is more expensive, the investor's implicit hope must be that the aggregate active portfolio will beat the passive implementation of the policy, by a sufficient margin to compensate both for the added cost and for the risk of failure. The margin is usually called the excess return or the value added.

The traditional presentation of active management results

Let's ignore the fact that traditional presentations are made against a sample (often exaggeratedly called a universe) of managers with a presumed similar benchmark. Let's also ignore the fact that some managers make such large, and changing, bets that it's tough to identify an appropriate benchmark for them. (Sharpe has designed a useful technique to identify an appropriate benchmark for the manager's activities in retrospect.[1]) Let's look only at the traditional comparison of the portfolio's return against whatever we have identified as the appropriate passive benchmark.

---

[1] Sharpe, William F. (1992, Winter). "Asset allocation: management style and performance measurement", *Journal of Portfolio Management.*

Here's what the traditional presentation looks like[2]:

Exhibit 1: Annualized value added for most recent periods ending on (some specified date)[3]

Length of period:	Quarter	1 year	3 years	5 years	10 years
Value added:	+47 bp	+80 bp	+317 bp	+181 bp	+52 bp

(annualized for periods greater than one year)
bp = basis points

Note that all the periods end on the same date. The most recent quarter's number is therefore implicitly included in all the other numbers. The single quarter excess return of 47 basis points would annualize to 189 basis points over a full year, so the three quarters before the most recent quarter must have been, on average, less successful than the most recent quarter, because they reduce the annualized value added. Similarly the two previous years must have been more successful than the most recent year, as the three-year average is substantially higher than the one-year number. The fourth and fifth years before the specified date must have been less successful than the most recent three. And the five earliest years may well have been years in which value was lost by active management, given that the 10-year number is so far below the five-year number.

Problems with the traditional presentation

One problem is that the implications for previous periods are all concealed by the common end-point. Far from being five numbers that typify the period, each number is embedded in the one that follows. The most recent quarter (much more successful than the 10-year average in the example) is embedded in all the other numbers.

We could unravel the chronology by listing the value added in each of the 40 quarters separately. But that starts to become very messy, and difficult to interpret. Several positive numbers, several negative: should the investor be happy or sad? As an investor myself, as well as having been for decades a consultant to institutional investors, ideally I'd like to see a presentation in which the interpretation leaps out at me: "Look, this is clearly good!", "Oh dear, this is clearly bad!" or "This is all over the place!" Well, there's no perfect answer. Statistics rarely tell the story clearly. But with trial and error, about 10 years ago, I found that looking at three particular dimensions of performance assisted my interpretation considerably. Let me share them with you.

Start with a picture

I prefer pictures to numbers. Most people find them easier to interpret. So my starting point is to suggest that we bring out from the physical or electronic warehouse the 40 quarterly reports since a manager was hired, and simply plot the value-added numbers reported in them. (There is no magic in quarterly numbers. These days numbers are typically reported monthly. The principles are the same.)

---

[2] This traditional presentation does not conform with the Global Investment Performance Standards ("GIPS") developed by the CFA Institute, for presentations by investment managers who claim compliance with GIPS. They require at least 10 annual returns (not annualized) for the manager, side by side with corresponding annual returns for the benchmark.

[3] This is a hypothetical example that is meant solely to illustrate how to interpret performance calculations. It is not meant to reflect any actual performance.

The result looks like this[4]:

Exhibit 2: Value added by manager since inception, over various rolling time periods

Excess return summary
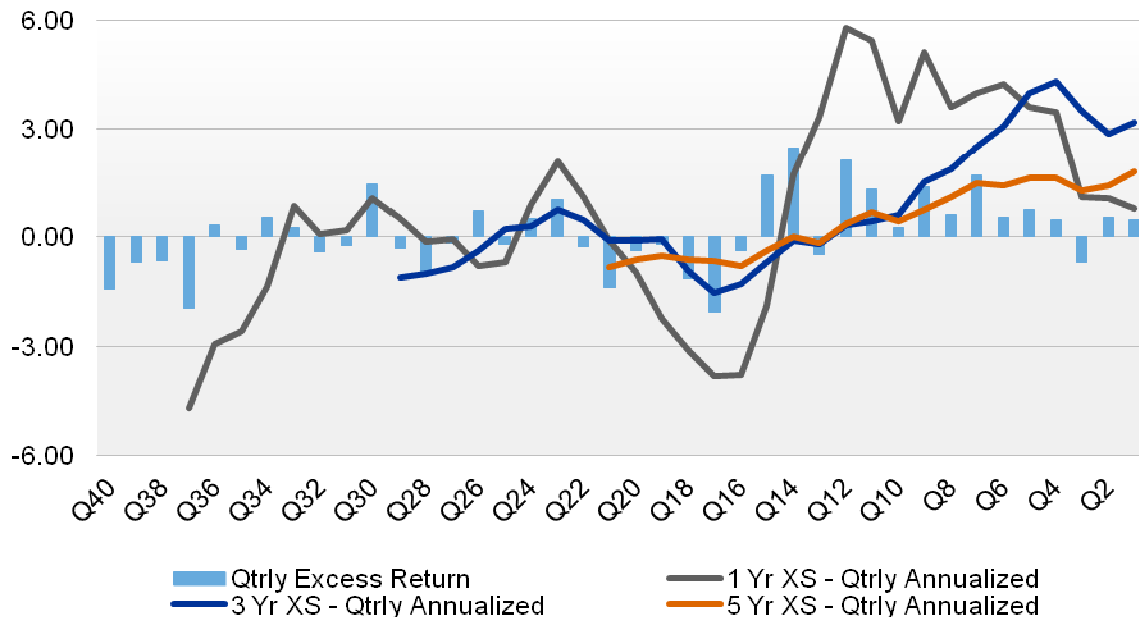
Now we see … what?

Well, one thing we see is why the most recent numbers are inadequate. They are simply the most recent plot points, the ones on the right-hand end of each sequence. No wonder they can't tell much of a story. How can they possibly be representative of the whole chart?

When I demonstrate this to institutional investment committees, it's often an "aha" moment. They suddenly realize that the traditional statistical basis on which they've evaluated the performance of managers (whether already hired or candidates for hire) is both flawed and easily improved.

They no longer want to see overlapping numbers, which are influenced considerably by the most recent period. They want the whole series.

But then, of course, they want the series condensed into a few representative numbers! Well, who doesn't? So, if the traditional presentation isn't representative, what is?

We don't have to reinvent the wheel. Statisticians typically present the mean and the standard deviation of a series of numbers as a quick summary of the whole set. We can do the same thing.

But first: which numbers to use?

Two sets of issues invariably arise. Which excess returns to use: gross or net of fees? And how should one compound the numbers?

---

[4] The numbers shown here are from a foreign country, and the data were provided for me by Russell analysts in that country, without identifying the managers. I know nothing about the managers other than what the analysis tells me. Again, the numbers are shown solely to illustrate how to interpret performance statistics and are not meant to represent anything else.

"You can't eat gross returns" is one way to remind the investor that only net returns are available. Yet, for this purpose, I prefer to use gross returns. Why? Clearly the onus is on me to justify my preference.

It's because of the question I want to use the numbers to answer. My question is: does the manager have skill at active management? For that, I want to see if the manager's returns beat a passive index. How much I pay the manager for that skill is an important question, but a different one. If the manager is capable of generating 100 bp per annum, that's a skill. It's worth my while to pay 50 bp for that skill, but not to pay 200 bp. An investor paying 50 bp is getting a good deal, another investor paying 200 bp isn't. But both investors can agree that the manager has skill, even if the second investor admits it through clenched teeth.

The other invariable issue relates to compounding numbers that are measured over different periods. This is done in many ways, without standardization. There is, however, one way that has some meaning, so I will use it. It is called the "geometric excess return", and it is calculated as:

(1 + actual annualized return)/(1 + benchmark annualized return) – 1.

This means it is the constant return, compounded annually, that connects the ending wealth generated by the actual portfolio with the ending wealth generated by benchmark returns (in the absence of cash flows).

A subsidiary question that often arises is: should one annualize the returns for periods shorter than a year? Traditional practice is not to do so. In the tables that follow, I will annualize them, simply to ensure that we are looking at annualized numbers over all time periods and don't have to play with the quarterly excess returns before comparing them with annualized returns over the other periods.

FIRST DIMENSION: DID THE MANAGER ADD VALUE?

This is the fundamental question.

Over the 10-year period shown in Exhibit 2, the (geometric) annualized excess return of the manager over the relevant benchmark was + 52 bp.

How much the manager was paid is a separate question, which I will not deal with.

Does the positive excess return demonstrate skill? Or was it just luck? Nothing will ever enable us to answer that question definitively. But there are other measures that may provide some insight.

SECOND DIMENSION: HOW VOLATILE WAS THE ADDED VALUE?

This is where the standard deviation comes in handy. Calculate the annualized standard deviation of the quarterly series. (Not the annual or longer-period series, because those contain overlapping quarters, so the numbers aren't independent of one another. In fact, using the monthly series is even better, because this ensures a greater number of data points and a more robust calculation.) The lower the standard deviation, the more regular (and therefore consistent) the excess return; the higher the standard deviation, the more volatile (and therefore the less consistent) the excess return.

In investment jargon, this standard deviation is called the "tracking error" of the excess return. The higher the tracking error, the less closely the excess return tracks its average. Now, it is quite possible for a manager to have a high tracking error, but still deliver a return stream that is less volatile than that of the market—and that would be a good thing. However, in the context of the active management assignment, other things being equal, the investor prefers a lower tracking error.

That doesn't mean the investor would not prefer to live with a higher tracking error, if that comes with a higher excess return. In fact, a derived measure, the "information ratio," is often calculated. It's the excess return divided by the tracking error. Interpreting the tracking error as a measure of risk, a comparison of information ratios answers the question: "Which manager better compensates me for the risk I'm taking with them?"[5]

THIRD DIMENSION: HOW CONSISTENTLY WAS VALUE ADDED?

I've found that, in practice, investors like the series in Exhibit 2 to be consistently positive. It's nothing technical—it's just a very human tendency to feel nervous when confronted with pluses and minuses at random. It doesn't matter if a series of pluses and minuses adds up, at the end, to an aggregate plus. Investors find it difficult to do a Rip Van Winkle and go to sleep for 20 years and hope to wake up to an aggregate plus. They examine ongoing results. That's why some consistency in the pluses is very desirable.

I use a very simple measure for this. What proportion of the time was each series positive? This is what the result looks like, for the manager in Exhibit 1:

Exhibit 3: Consistency of delivering added value since inception, over rolling time periods
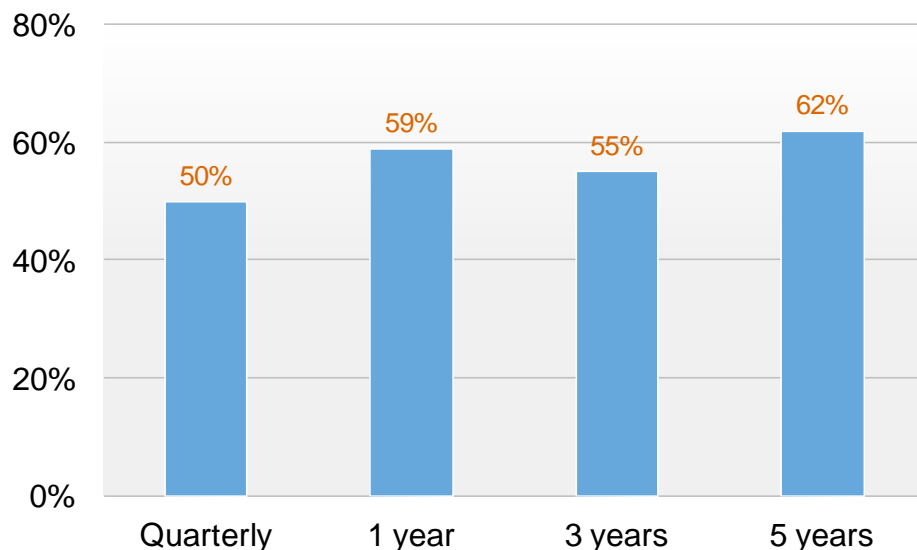
Source: Russell Investments

Intuitively, these feel very close to 50%, a number that suggests that randomness or luck is at play.

It's instructive to ask why luck would play a role at all, and what sort of pattern a skilled manager would be associated with.

---

[5] The tracking error is sometimes interpreted as a good measure of the size of the bets the manager is taking. Wrong. For example, consider the extreme case of a manager with an excess return and zero tracking error. Zero tracking error can't mean no bets, otherwise how would the excess return arise? It means consistent excess return, not zero bets. That's because the tracking error is measured around the average excess return. If it were measured around zero excess return, then it would be a better (ex-post) measure of the size of the bets taken. Assuming no correlation between excess returns and tracking error, a quick-and-dirty measure of the bet size would be:

Bet size $\approx \sqrt{[(\text{excess return})^2 + (\text{tracking error})^2]}$

The use of this equation is beyond the scope of this paper.

Think of what an active manager does. The manager selects securities that are considered undervalued, relative to the rest of the market. What's the test of whether they are undervalued? It's for their value to rise, relative to the rest of the market. That means that other active investors (who have established the current values of the chosen securities) must change their minds and trade up the values of this manager's securities. Whether other managers will change their minds, and in particular whether they will change their minds in the next quarter, are not within the manager's control. If the manager is indeed skilled at identifying undervalued securities, then it's likely that, for most of the manager's selections, other investors will eventually come round to the manager's viewpoint and will trade up the securities. But when that will happen – the next quarter, the next year, or whenever – is typically a matter of chance. That's why, even with a skilled manager, short-term value added can be negative at times (and indeed, success ratios for any manager—skilled or not—typically show up in the 50% range when measured over quarterly periods). But as the time horizon lengthens, skill shows up as a rising success ratio. Equally, poor judgment shows up as a declining success ratio as the time horizon lengthens.

And success ratios that stay close to 50% regardless of the length of the time horizon? It's not unreasonable to question whether the results reveal any skill at all, or just the random effects of chance -- as shown in Exhibit 3.

Is there a natural period of time for skill to start showing up? I think there is. It's the inverse of the manager's turnover ratio. For example, if a manager's turnover ratio is 25% per annum (p.a.), the inverse is four years. If the turnover ratio averages 50%, the inverse is two years…and so on.

Why is this relevant? Well, if the turnover averages 50%, it means that half the securities in the portfolio are typically sold and replaced every year. The average security is held for two years. Within that length of time, the manager expects skilled selection to reveal itself. That doesn't mean that, over two years, value is *bound* to be added. But I think it does mean that, over *many* two-year periods, skill ought to show up. I also think it's unfair to judge such a manager over periods shorter than two years, because the manager gives the securities two years (on average).

Nevertheless, I find in practice that investors like to look at success ratios over all periods, starting with quarterly, even when they know that quarterly periods are dominated by chance.

More examples

Exhibit 4 shows traditional performance reports for two managers.

Exhibit 4: Annualized value added for most recent periods ending on (some specified date)

Length of period:	Quarter	1 year	3 years	5 years	10 years
Manager A:	-520 bp	-367 bp	+54 bp	+266 bp	+475 bp
Manager B;	-904 bp	-368 bp	-335 bp	-327 bp	-149 bp

Manager B looks awful, at least recently. But perhaps longer history showed some success. Manager A looks up and down (or rather, down and up!). Where's the consistency?

Plot their series as in Exhibit 2 and calculate their three dimensions of excess return, and the results look a little clearer – and perhaps different from one's first impressions.

Here are their average excess returns and tracking errors:

Exhibit 5: Annualized excess return and tracking error for the 10 years ending on (some specified date)

Length of period:	Excess return	Tracking error
Manager A:	+475 bp	512 bp
Manager B;	-149 bp	271 bp

And here are their success ratios:

Exhibit 6: Success ratios over rolling periods of different length in the 10 years ending on (some specified date)

Length of period:	Quarter	1 year	3 years	5 years
Manager A:	65%	87%	96%	100%
Manager B;	43%	41%	17%	10%

These statistics confirm that Manager B was a disaster – though that might not have been apparent over short periods. Certainly the most recent quarter was terrible, but the 43% and 41% success ratios in Exhibit 6 suggest that at times when the most recent quarter was positive (and nearly half of them were), the traditional presentation (which is always overly influenced by the most recent quarter) probably looked pretty good.

There's not much to read into the tracking error stat, in the absence of a sample of tracking errors for managers with that type of mandate.

Manager A now looks outstanding, though the tracking error is higher than for Manager B. But the value added is terrific, and the success ratios are absolutely exemplary. This is what success looks like, over the long term. But what on earth happened in the most recent quarter? Clearly that's an aberration. It's roughly two standard deviations (two tracking errors) below the average long-term added value.

That's a big discrepancy. It suggests that we want to talk to this manager, to find out whether his or her approach to active management has changed, and to ask related questions. The numbers can't answer those questions. But they can point out aberrations that prompt the need for questions and an interview. Whatever comes of it, it's clear that the traditional presentation distorted this manager's numbers unfavorably, when one is armed with the additional insights that this 3-D analysis reveals. (I call this analysis 3-D, simply for my amusement.)

Obvious (but important!) limitations of this analysis

One reaction I have encountered, when showing an institutional investment committee this three-part analysis of its fund's managers, is excessive enthusiasm. For example, it has been hailed as a decision tool: an indicator of whether or not to retain a manager, and indeed when to do so. The analysis is nothing of the sort.

It is simply a more evocative way of presenting past performance than the traditional presentation. In this regard, it's useful. It's as if the available data were a list of words: the traditional presentation shows some of them, but the 3-D analysis rearranges all the available words into sentences and paragraphs. The available data tell a story, and the 3-D analysis helps the story to emerge, while the traditional presentation often hides the story. That's good, but that's all.

And one reason it's good to look at three dimensions instead of just one is that any single measure can be misleading at times. For example, the "success ratio" analysis would not work well if a manager has an approach that tends to produce frequent small gains and infrequent large losses (or vice versa). Multiple measures of success tend to build a more complete picture.

There is nothing at all in the 3-D analysis that explains *why* the story has emerged as it has. It's simply some facts, not an explanation. There's no way to know even whether the same investment management team, or active management approach, has been in effect throughout the period. Only further investigation – talking to the manager, for example – can do that.

And as for the notion that 3-D analysis constitutes a timing tool for hiring or firing – that's even less valid. There's no point at which 3-D indicates: "Now's the time to hire" or "Now's the time to fire." There is no statistical significance attached to the numbers I look at.

Nor does 3-D help to predict the future. It only looks at the past.

One final downer: the 3-D numbers are so far from traditional that they are simply not routinely presented. They require some effort to calculate, as we have seen.

Two other applications of 3-D

There's nothing that restricts 3-D to evaluating active managers. I've found two other applications useful.

One is to test institutional committees on their own manager selection. Compare the returns for their total fund with the corresponding returns for passive implementation of their selected asset allocation policy. The differentials constitute a commentary on their own skill in selecting managers. (And looking at total fund returns rather than asset class returns or single manager returns makes the analysis more robust, revealing more of the skill signal and eliminating more of the random noise that all returns contain.)

Committees are much more willing to evaluate managers than they are to turn the mirror on themselves. (Aren't we all?) Some will feel thoroughly vindicated by rising success ratios as the rolling time periods lengthen. Others, with falling success ratios, must face the interpretation that it is not managers who constitute the problem, but their own processes of evaluating, hiring and firing managers that have failed.

The other application is in the context of asset allocation.

Here we're typically looking at something like the equity/debt ratio that is most suitable for a fund, and trading off the expected long-term superiority of equities against the expected short-term volatility reduction associated with debt. Typically we think that the superiority of equities over debt becomes more apparent, the longer the time horizon. Indeed, most of us typically think that we'd far rather bet on equities outperforming debt over the long term, than on our ability to pick a skillful active manager – or indeed on the ability of a good active manager to add value from active bets.

The test is simple:

- Compare equity index returns with debt index returns over various moving periods,
- Define success as equities outperforming debt, and
- Measure the consistency of success, the average amount of success, and the risk involved.

Over very long periods, you'll typically find the equity success ratio rising as the averaging period lengthens. The amount of success defines the equity risk premium over the period. You'll find a high tracking error, meaning that the equity premium is highly volatile.

Now compare the results with those of a successful active manager. You may be surprised which turned out to be more reliable!

Acknowledgements

I thank Shailesh Kshatriya for valuable contributions to this paper.

For more information:

Call Russell at **800-426-8506** or
visit **www.russell.com/institutional**

Important information