The perils of a CPI benchmark

By Amira Abbas, Research Scientist, STANLIB Multi-Manager

A discussion of the pitfalls of using one benchmark in particular, CPI (the consumer price index), with regard to performance evaluation.

Evaluating the performance of an asset manager is often subjective. Broadly speaking, there are three elements to consider:

- Measurement – choosing an appropriate benchmark
- Attribution – determining which asset classes or instruments contribute to relative performance
- Appraisal – understanding the meaning of the performance

All come with their own complexities, regardless of how a manager structures their fund.

In this article I discuss the pitfalls of using one benchmark in particular, CPI (the consumer price index), with regard to performance evaluation. Instead of addressing how performance evaluation should be done in this context, I will instead focus on the important dimensions to consider.

Why is CPI used as a benchmark?

Despite its limitations, CPI is widely used as a benchmark, and with good reason. In a goal-based world, an absolute return (nominal or real) is necessary if you are saving towards a goal, as you need a mechanism to discount your investments and liabilities. CPI plus an additional percentage represents a real return over inflation and makes sense to all sorts of investors.

Contrast this with a target of 2% alpha over an index other than CPI, say an equities index. This gives no insight into the relative value of money. The fund could have achieved its objective of say 5% while the index returned 3%. If inflation over the period was 6%, the value of money invested and its purchasing power decreased because it's eroded by inflation.

What makes a good benchmark?

A benchmark should serve as a point of reference to which other things may be meaningfully compared.

*A **good** benchmark is one that is fully consistent with a manager's investment process, style and philosophy. This consistency makes it easier to evaluate the skill of a manager, and their ability to exploit perceived opportunities by taking "off-benchmark" bets that translate into alpha.*

An appropriate benchmark serves as a sanity check to ensure a manager follows the style and philosophy they claim to follow.

Benchmarks should communicate information about the manager's investable universe and provide an indication of acceptable levels of risk versus return. To do this they need to conform to the characteristics of good benchmarks:

- Investable – it should be possible to replicate and hold the benchmark, i.e. the weights and securities in the benchmark should be identifiable and available for investment

- Appropriate – the benchmark should be consistent with investment style and reflective of the manager's investment opinions

- Accountable – the benchmark chosen signifies the manager accepts ownership of the constituents and is held accountable for significant deviations[1]

Why is CPI a bad benchmark?

CPI is not investable

Asset managers cannot directly invest in CPI and achieve its return – unlike an equity index, for example, which may be closely replicated by purchasing the index's underlying stocks. Even inflation-linked bonds do not represent investing in CPI.

CPI does not provide any information about a manager's philosophy or investment style

It may appear as if a manager is risky by taking large off-benchmark bets (assumed from a large tracking error relative to CPI), but this is merely because most asset class returns are volatile relative to CPI over time. This makes it difficult to evaluate a manager's performance.

How do we evaluate a fund with a CPI benchmark?

From the get go, let us be clear that we are going to separate manager performance evaluation from fund performance. We will come back to how to evaluate manager performance when the benchmark given is CPI. Investors will often question why a fund does not beat its benchmark every year. They fail to recognise that returns (both nominal and active, that is relative to the benchmark) are inherently uncertain, which is the very definition of risk in investments. There are however two important dimensions that aid with the understanding of this risk.

The first is the dispersion of the returns (which can be measured around its average value in the case of nominal returns, or around a benchmark in the case of active returns). The second, is how the certainty of the average increases with the sample size, that is the average value of the returns becomes more certain as we extend the period over which we measure returns (this is often misstated as time diversification). Yes, the above includes some implicit assumptions about the processes generating the returns, but let us ignore that complexity for the sake of not getting overly complex.

While the above uncertainty exists even for great benchmarks, things get more complex when we consider CPI objectives as benchmarks, since asset class returns are generally volatile (uncertain) by comparison.

So how do we choose an appropriate time period for evaluation, and how certain can we be that we will achieve the CPI objective over that time period? These questions are difficult to answer and involve a lot of subjectivity, even when conducting quantitative analyses.

A simple example

Consider a SA equity-only fund, with a benchmark of CPI+7%. We will look at excess returns and examine the complexities that arise in assessing the performance of the fund.

Probabilities versus active returns and holding periods

Active returns (less than or equal to)	Holding period			
	1 year	3 years	5 years	7 years
-30%	2%	<1%	<1%	<1%
-20%	7%	<1%	<1%	<1%
-10%	22%	9%	4%	2%
0%	46%	44%	42%	41%
10%	72%	85%	91%	94%
20%	89%	98%	99%	99%
30%	97%	99%	>99%	>99%
40%	99%	>99%	>99%	>99%
50%	>99%	>99%	>99%	>99%

The table above uses historical South African equity asset class returns relative to CPI+7%. If we assume that future returns can be parameterised based on these historical returns we can estimate the probabilities of certain outcomes over different time horizons.

Consider the first row of the table where the probability of underperforming CPI+7% by 20% or more, is about 7% if we look at a one year holding period, but less than 1% for three, five, and seven year holding periods.

There are important take-aways from this table.

Notice how active returns (annualised) fall into a narrower band as the holding periods increase (commonly referred to as the funnel of doubt). The probabilities are contained between active returns of approximately -10% to 20% for 7 year holding periods as opposed to -30% to 40% for 1 year holding periods. This confirms our previous contention that the sample average becomes more certain as the sample size increases, that is the return becomes more certain as the holding period increases. We often hear "if you're investing in equities, you need to be invested for a longer time horizon due to higher volatility of returns", so this makes intuitive sense.

Given that we can expect the results to become more certain as we increase the time frame over which we do the analysis, how do we decide on how long is enough? Surely waiting forever is not an option. In fact, we probably want to wait as little time as possible, for a number of different reasons. We should therefore understand that we will need to compromise between waiting too long to be certain as the information will become worthless, and waiting too little and being very uncertain. There are three variables that we can flex to help us address this.

1. The certainty of achieving the objective

How certain do we want to be (ex-ante) of achieving the objective? Remember that in financial markets (as in life) nothing is certain (not even death and taxes, as some countries have zero taxes, and pond scum are immortal).

Say we would like to be 60% sure that we outperform the CPI objective. Using Table 1, we can work backwards

to see what time period corresponds to a 60% chance of achieving an alpha of 0% or more (i.e. a 40% chance of achieving 0% alpha or less). Looking at the row containing 0% alpha, this corresponds closely to a seven year holding period. Therefore, considering a fund's performance over a seven-year period, would only get us to being 60% sure that the fund would achieve this benchmark. Unfortunately, that also means that there is still a 40% chance that the fund will underperform this benchmark (a very likely event).

So what can we conclude ex-post if the fund underperforms or outperforms the benchmark? Well, very little, especially when the probabilities are so high in both cases. Ideally, you want an event to have a very low probability to conclude that it is unlikely to have occurred by chance. Unfortunately, unlikely does not imply that it can't occur, and you could still get to the wrong conclusion.

Also, because these probabilities are very close to 50%, extending the time period even longer will not help much either. These probabilities are close to 50% because the historical returns were close to CPI+7%. We will therefore need to flex a different parameter if we wanted to increase the certainty of achieving the objective.

2. The time horizon (holding period of the analysis)

Suppose we do not know how certain we want to be, but we know over what time frame we would like to do the assessment. Perhaps the manager has provided guidance that they aim to achieve the objective over three year rolling periods. In this case we can use our model to see what probability is associated with the three year period. Again, looking at the row containing 0% alpha, and the 3 year column, we see that underperforming the benchmark translates into a 44% probability (a 56% probability of outperformance).

This is again close to 50% for exactly the same reasons highlighted above. You will notice that these two variables are closely related, that is increasing the certainty requires increasing the holding period and vice versa. At this point, things may be looking a little bleak, but we have one more variable that we can flex, and this one will come to the rescue.

3. The CPI objective (or benchmark)

The third variable that we can vary is the CPI objective itself, or equivalently, the alpha sought. In the above example, we chose an objective that was close to the historical average return of the asset class. We should therefore expect the probabilities of outperforming or underperforming this to be close to 50%, by definition of our model of future returns.

If we instead consider a benchmark of CPI+5%, the probability of underperformance drops. Now the dimension of time makes a much bigger difference. If we again consider Table 1 above, at 0% alpha, the probabilities of underperformance are 46%, 44%, 42% and 41% for one, three, five, and seven year holding periods respectively. With a revised objective of CPI+5%, these probabilities would drop to 42%, 36%, 32% and 29% for the same respective periods (not shown in Table 1). The benefits of a longer time period become more pronounced as expected.

There is an important compromise that is happening here, that we shouldn't lose sight of. By lowering the target we are measuring against, we are improving the chance of achieving (or more accurately exceeding) it. We are not changing the expected outcome, and we should be careful to not confuse these two issues.

This is in some ways analogous to making predictions or guessing the value of quantities you don't know. One way of improving your success rates, is to make your prediction or guess less precise – perhaps a wider range.

An analogy will help clarify this point.

> *I ask you to estimate the circumference of the moon in kilometres, and to be 50% confident in your answer. If I asked you to estimate many different things at this level of confidence, you should expect to get approximately half of them right, and half of them wrong. You may have no idea what the circumference of the moon is so you probably want a fairly large range for your estimate for example between 8 000km and 15 000km.*

> *If I ask you to be 99% confident in your answer, you would aim for a much wider range, say between 5 000km and 20 000km. Incidentally, the circumference of the moon is approximately 10 921km.*

Back to manager performance evaluation

I promised to come back to this and provide some guidance about how you would do this in the context of CPI+ benchmarks. In a nutshell, you throw out the CPI+ benchmark, and substitute it with something more appropriate. In some cases this will be fairly straightforward, for instance use an appropriate equity index for an equity mandate. In other cases it may be a little more complex, such as what do you use for a balanced or high equity multi-asset class fund (a composite of various indices may be appropriate in these cases). There are many tools and techniques to help with this exercise, and tracking error (or equivalently, the r-squared from a linear regression) is a good starting point.

Conclusion

Performance evaluation can be complex at the best of times, and downright impossible at the worst of times. Performance evaluation versus traditional benchmarks has its complexities, but may at least provide insight into the skill of an asset manager.

CPI objectives on the other hand, make economic and intuitive sense, but introduce a range of unique complexities, making performance evaluation impossible. It is important to understand the uncertainty that arises when faced with these benchmarks, and the variables that can be flexed to reduce this uncertainty. The confidence (certainty) in the results, and holding period for the analysis are two such variables. A third is the objective itself (or the level of relative performance sought).

Unfortunately, this does little to help in evaluating the performance of the underlying manager, but there are a range of tools and techniques that can assist in this exercise.