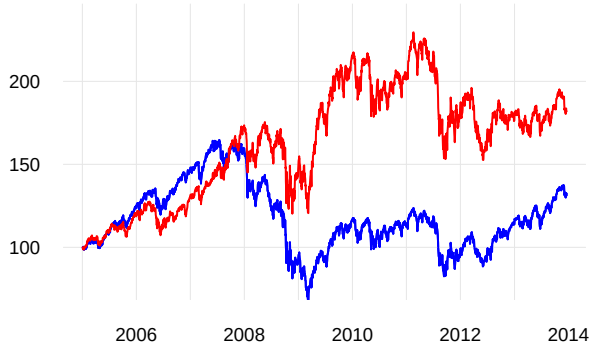
Two pitfalls in comparing financial time-series

Enrico Schumann
es@enricoschumann.net

20 December 2013

It is sometimes quipped that few decisions determine the success of a portfolio manager as much as the choice of a benchmark. Indeed, first – and later – impressions are often driven by graphical displays such as the following one.



Such graphics are ubiquitous in financial reporting whenever it comes to comparing a fund with a benchmark or with another manager. But such displays can be misguiding – give a false impression – in at least two ways,[1] as we shall discuss in this note.

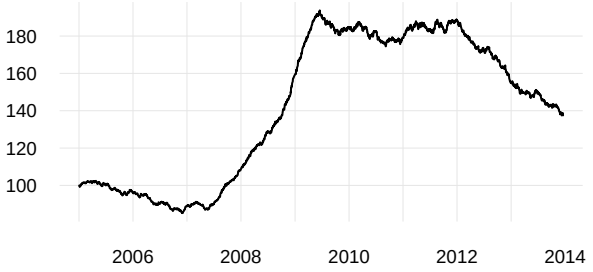
Suppose the above graphic shows how 100 euros would have evolved with managers BLUE and RED. Then see if you can agree with the following two assertions:

1. RED's return was higher than that of BLUE.
2. RED's portfolio was more volatile than that of BLUE.

Agree? Then let us discuss both statements.

RED's return was higher than that of BLUE

That is undoubtedly true. But often we also care about *when* one manager performs better than another, and that is more difficult to gauge from the chart. In fact, when we care about who performed better, we should directly look at the difference. In other words, if we care about the *difference* between two things, then we should plot this difference. Here it is:



The chart shows the ratio of RED to BLUE. Thus, when RED performs better than BLUE, the line is rising; if RED performs worse, the line goes down.

From the chart we can read that RED performed better by about 40%. That is, an investment with RED would have led to a final wealth that was 40% higher than that of an investment with BLUE. But we can also see that RED's advantage originated almost entirely in the years of 2008–2009, when equity markets fell. In other periods RED actually did worse than BLUE, notably since the beginning of 2012.

> If you care about the difference between two things, then plot it. (That is, plot one thing, not two.)

RED's portfolio was more volatile than that of BLUE

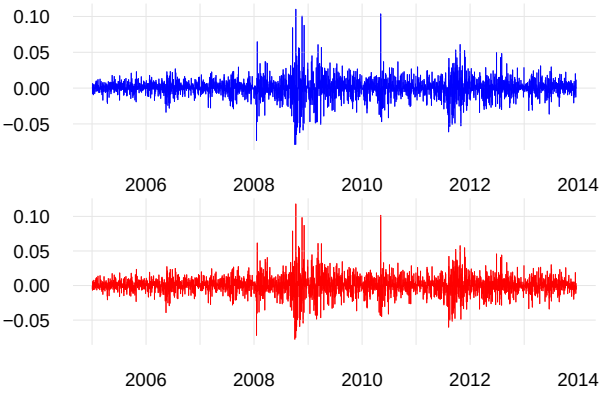
You probably suspect a trick question, and indeed, both time-series are equally volatile. That is, the returns of both series have the same standard deviation of about 1.4 % per day or 23 % per year.

When we speak of volatility, we typically mean relative changes. But in the picture, we look at absolute changes. The impression that RED is more volatile follows from the fact that absolute changes for RED are greater than for BLUE.

The same holds true for related measures such as drawdown. Granted, RED had a smaller drawdown during 2008–2009, but what about the summer of 2011? Again, you suspect a trick question, and, unsurprisingly, both series have the same maximum drawdown in 2011 (about 33 %).

We can make that even clearer by artificially improving RED's performance; see the next chart. In that chart both series still have the same daily volatility, and the maximum drawdown in 2011 is still the same for both series. But RED clearly appears more volatile.

Indeed, if we really wanted to look at volatility, we should do so. As already admitted, there is no difference.



> It is difficult to gauge differences in price variability (as measured through return volatility or drawdown) when price levels are very different.

[1] There are, of course, the obvious shenanigans such as abusing the aspect ratio. But the two pitfalls I describe are typically not deliberate attempts to fool someone.