A note on 1/N and minimum-variance portfolios, and the fact that significance tests do harm (and no good) in financial decision-making

Enrico Schumann
es [AT] enricoschumann.net

28 November 2012

Formal tests of statistical significance (in the sense of null hypothesis tests) are of no value for when it comes to financial decision-making. The reasons why I think so are not original or new; they have been persuasively described for instance in McCloskey and Ziliak (1996) or, more relevant to our topic, in Armstrong (2007a,b). In fact, there have been critics of statistical testing in various disciplines, for instance in psychology or in political science. Examples of such discussions can be found in Cohen (1994), Gigerenzer (2004) or Gill (1999).

I will explain this objection through a concrete example, the paper on 1/N-portfolios by DeMiguel et al. (2009). This is not because I have anything against this paper in particular, but because the conclusion that many seem to have taken from it is that 'you can't beat 1/N' – a conclusion with which I disagree. In fact, I will argue that the authors of the 1/N paper provide further evidence that the long-only minimum-variance portfolio is a reasonable investment strategy that is preferable to 1/N.

See Gilli and Schumann (2011) for references to empirical studies on long-only minimum-variance portfolios.

So here are some thoughts and observations on the claim that long-only minimum-variance (MV) does not do better than 1/N:

(i) We should never expect a single strategy to be better than another strategy in every dataset, at all times. Even a random portfolio will have its day. This is the nature of finance (and any other field that involves chance).

(ii) DeMiguel et al. (2009) compare 1/N with long-only MV for their empirical datasets, and they find the following monthly Sharpe ratios. (They only report results for six datasets. I give 4 decimals because DeMiguel et al., 2009 do; see their Table 3.) If the difference is positive then 1/N performed better than MV.

1/N	0.1876	0.1353	0.1277	0.2240	0.1623	0.1753
MV	0.0834	0.1425	0.1501	0.2493	0.1546	0.3580
difference	0.1042	-0.0072	-0.0224	-0.0253	0.0077	-0.1827

So in the tests of DeMiguel et al. (2009), the long-only MV portfolio has a higher Sharpe ratio than 1/N in 4 out of 6 datasets. The authors also acknowledge this good performance, albeit with a caveat that 'the differences are statistically significant only for [one] dataset' (p. 19). This reference to statistical significance is not helpful. A number of studies before DeMiguel et al. (2009) have found evidence that the long-only MV portfolio performs well when judged by out-of-sample performance (eg, Board and Sutcliffe, 1994, Chan et al., 1999, Clarke et al., 2006, Blitz and van Vliet, 2007). So the authors actually provide further evidence for this fact.

(iii) There is no discussion in DeMiguel et al. (2009) of the size of the differences, and in particular of the power of the statistical tests used. Already a small change in annualised Sharpe ratio can be meaningful: assume an equity portfolio returned on average 10% per year above the risk-free rate, with a volatility of 20%; the portfolio's Sharpe ratio would be 0.5. If the portfolio had a mean return of 11%, we would have a Sharpe ratio of 0.55. So a small difference (about 1.5 bp on a monthly scale) can already be meaningful. Of course we have sampling error, and we cannot know whether this advantage is for real. But consider the question of cost: what would be lost if we chose MV instead of 1/N? Under the null hypothesis, both portfolios are the same, so the worst case (as defined by our statistical test) is that while we think that we get a superior portfolio, we merely get a draw from the same return distribution. But apart from noise, we will not be worse off. We may argue that these portfolios are not equal, eg, when it comes to diversification, but that is not a fair objection: if we had wanted diversification, we should have written it into our objective function.

A careful statistical analysis would also require to discuss the power of the tests that were used; I conjecture that it is well below 50% for a small difference such as 0.02 on a monthly scale. Ironically, this implies that we would need a sample in which the difference between the strategies is, by chance, larger than the true difference (ie, a sample not representative of the actual difference between the strategies) to reveal the advantage of a strategy; see Schmidt (1996).

To be more to the point: the question that we ask is 'Is MV in general so much better than 1/N that it warrants the higher effort?', or more concretely 'Is the Sharpe ratio (or some other performance score) in general so much higher that MV is worth the effort?'. The a significance test's outcome is either

> 'It is *unlikely* that we would have obtained the given (or a greater) difference under the hypothesis that the difference is zero.' (small *p*-value)

or

> 'It is *likely* that we would have obtained the given (or a greater) difference under the hypothesis that the difference is zero.' (large *p*-value)

But we are not interested in a zero difference, so neither outcome answers our question.

The only way to demonstrate that a result is robust is to replicate it (Cohen, 1994, Armstrong, 2007a), and that is exactly what DeMiguel et al. (2009) have done: they have produced yet another study that documents the favourable properties of MV. It may not have been the authors' intention, but they have actually provided more support for MV.

References

J. Scott Armstrong. Significance tests harm progress in forecasting. *International Journal of Forecasting*, 23(2):321–327, 2007a.

J. Scott Armstrong. Statistical significance tests are unnecessary even when properly done and properly interpreted: Reply to commentaries. *International Journal of Forecasting*, 23(2):335–336, 2007b.

David C . Blitz and Pim van Vliet. The Volatility Effect. *Journal of Portfolio Management*, 34(1):120–113, 2007.

John L.G. Board and Charles M.S. Sutcliffe. Estimation methods in portfolio selection and the effectiveness of short sales restrictions: UK evidence. *Management Science*, 40(4):516–534, 1994.

Louis K. C. Chan, Jason Karceski, and Josef Lakonishok. On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies*, 12(5):937–974, 1999.

Roger Clarke, Harindra de Silva, and Steven Thorley. Minimum-Variance Portfolios in the U.S. Equity Market. *Journal of Portfolio Management*, 33(1):10–24, 2006.

Jacob Cohen. The earth is round ($p < .05$). *American Psychologist*, 49(12):997–1003, 1994.

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953, 2009.

Gerd Gigerenzer. Mindless statistics. *Journal of Socio-Economics*, 33(5):587–606, 2004.

Jeff Gill. The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly*, 52(3):647–674, 1999.

Manfred Gilli and Enrico Schumann. Risk–reward optimisation for long-run investors: an empirical analysis. *European Actuarial Journal*, 1(1):303–327, Supplement 2 2011.

Deirdre N. McCloskey and Stephen T. Ziliak. The Standard Error of Regressions. *Journal of Economic Literature*, 34(1):97–114, 1996.

Frank L. Schmidt. Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers. *Psychological Methods*, 1 (2):115–129, 1996.