Take-the-Best in portfolio selection,

or

The relevance of the assets' covariation for constructing portfolios,

or

The simplest way to construct the minimum-variance portfolio

Enrico Schumann
es [AT] enricoschumann.net

6 February 2013 (notes updated on 18 February 2013)

This is a draft; please do not quote it. Comments and corrections are welcome. Updates and code are available from `http://enricoschumann.net`

Many thanks go to Manfred Gilli for comments and suggestions.

A well-known result in portfolio optimisation states that as the number of assets in a portfolio grows, the variance of portfolio return approaches the average covariance between the included assets (eg, Elton and Gruber, 1977). In the language of financial economists, idiosyncratic risk can be diversified away; eventually only systematic risk remains.[†]

While this result may be as true as a mathematically-derived statement can be, its relevance for portfolio management is less clear. If it is taken to mean that diversification matters for portfolio construction, it is certainly valid. Yet if it is to mean that the portfolio constructor should care only about covariation – that is, the *joint* movement of prices –, then the result leads to a bad prescription for decision making. More specifically, the result should not be taken as a justification to emphasise forecasting correlations.

Covariation is a function of an asset's marginal distribution and its comovement with other assets. Which part of the asset's variation is thought to be marginal, and which part is thought to be caused by comovement depends on a model-based split, and such models need calibration, typically based on past data.

But we care little about the past; rather, we have to forecast risk.[†] If we keep the distinction between idiosyncractic and systematic risk, we need to forecast correlations and marginal risk; for portfolio risk we then aggregate these quantities. Such forecasts are crucial in portfolio optimisation. An algorithm will typically sweep through the space of possible portfolios, and in each step look at the expected risk of the current portfolio. If the risk forecast is not accurate, then the optimisation cannot help. In fact, it will do harm rather than good (Michaud, 1989).

I will argue in this note that drawing too much inspiration from the above-stated result is not helpful in forecasting portfolio risk. The reason is an asymmetry: if an asset is risky when looked at in isolation, then it *could* become less

[†] Throughout this essay, I will use the word *risk* interchangeably with *return variation* (or volatility, or return variance). This is clearly wrong: risk is the potential loss of a portfolio; that is, risk arises from trading decisions. Return variation is a statistical property and may or may not be correlated with risk (just think of a pegged currency). Nevertheless: using *risk* this way is (i) in line with the financial economics literature and (ii) not unreasonable for equity portfolios; and (iii) *risk* as a word, in a literal sense, is just so much better that *volatility*.

[†] Apart from forecasting, there is a second problem. Few people or even institutions hold portfolios of thousands of assets. How long would it take to 'diversify away' the peculiarities of marginal distributions? A number of papers in the 1970s tried to answer this question by formulating the problem as follows. We start with the maximum–expected-return portfolio, which consists of just one asset. We are not certain about our forecast, so we add another asset to the portfolio. This will necessarily reduce expected return, but it will probably also lower portfolio variance. Thus, there should be trade-off between expected return foregone and reduced volatility. This is an elegant and clear story, but practically of no use since we have little evidence that returns can be forecast. But we can typically forecast volatility.

risky in a portfolio because of low correlation. But if an asset is not risky on its own, it *cannot* become riskier than implied by its marginal distribution. A forecasting scheme should take into account this asymmetry.

An alternative (and much simpler) description is to say that assets have several desirable properties, such as 'looks good on its own' or 'looks good in relation to other assets'. The goal of this note is to show that only considering 'looks good on its own' provides so much information that including 'looks good in relation to other assets' does not add much – because we have to make forecasts of these properties, and we are likely to have forecast errors when doing so.

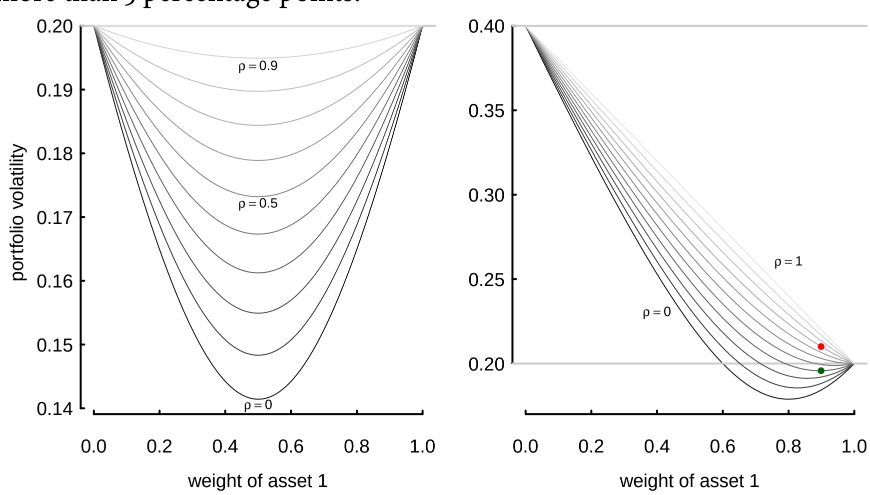
This point of view is reminiscent of the Take-the-best heuristic in psychology (Gigerenzer and Goldstein, 1999). Take-the-best prescribes to base a decision solely the most important cue. The rule will be successful in so-called non-compensatory environments: if we get the most important dimension right, other dimensions add no incremental information. I will argue in this note that portfolio optimisation (more specifically, minimising portfolio risk[†]) may well represent such a non-compensatory environment.

[†] A large number of studies document that caring most about portfolio risk, and not return, works well empirically. See Gilli and Schumann (2011) for references and discussion.

Intuition

Imagine a long-only equity manager who wishes to reduce the return variation of his portfolio. For intuition, suppose there are only two equities, both having a volatility of 20%. The following figures illustrate the effect of correlation on portfolio volatility.

For possible correlations between zero and unity, the left-hand side of the picture shows portfolio volatility given a specific invesment in asset 1 (the weight of asset 2 is one minus the first weight). Unsurprisingly, portfolio volatility is minimised by putting 50% into both assets. With zero correlation, we would reduce risk by more than 5 percentage points.



Practically, we do not have zero correlations, and not all assets have equal volatility. So when we set the second asset's volatility to 40%, we get a different picture, as shown on the right-hand side. We see that independently of correlation, we will put almost our total budget into the first asset. As a concrete example, with a low correlation of 0.3, we could reduce portfolio risk by less than half a percentage point. This specific point is indicated by the green dot • in the right-hand side graphic.

As long as there is substantial variation in the cross-section (which empirically is the case), we would not lose much even if we were able to accurately forecast covariation. So we have a first argument for ignoring correlation and selecting assets based on 'looks good on its own'.

And there is a second, even stronger, argument. What happens if we make a mistake in forecasting correlation? Again, look at the right-hand side picture above. If the true correlation is 0.6, the optimal allocation would be 100% in asset 1. If we forecast correlation to be 0.3 instead we would (erroneously) allocate wealth to the second asset, and not get a portfolio risk of 20%, but of about 21% (indicated by the red dot •). Thus, forecast error will leave us worse off – which we could have avoided by only looking at the individual volatilities.

A simulation study

In this section I will present the results of a small simulation study. The aim is to compute the long-only minimum-variance portfolio from a number of assets.

Computing is done with R (R Development Core Team, 2013). The minimum-variance portfolio is computed through quadratic programming with the `quadprog` package (Turlach and Weingessel, 2011). This manuscript is written with Sweave (Leisch, 2002); the code is available from http://enricoschumann.net .

Specifically, there are 100 assets from which we can choose; marginal volatilities range from 20 to 40% percent. The goal is to compute the long-only minimum-variance portfolio of these assets under the restriction that all weights are between zero and 5%.

We will test two methodologies: *classic*, for which we compute the variance–covariance matrix, and *sort*, for which we simply pick those 20 assets with the lowest marginal volatility, and put 5% into each.

We use random data sets with constant pairwise correlations ρ, which we vary between 0.1 and 0.9. For each value of ρ, for a single simulation, we (i) randomly draw a dataset of 200 observations, (ii) compute the long-only minimum-variance portfolio via methods *classic* and *sort*, and (iii) and compare the results of both methods. Altogether, we repeat this procedure 1000 times for each value of ρ.

For (ii), the input parameter is the sample variance–covariance matrix (*sort* will only use the matrix's main diagonal). For (iii), we compute two quantities: first, the expected – or rather 'hoped-for' – advantage of the *classic* portfolio over the *sort* portfolio, based on the specific sample:

Even better would have been to specify true parameters and then randomly add errors; this would serve as a reminder that practically we should not *estimate* the required quantities, but should *forecast* them.

$$\text{expected difference} \quad = \quad \text{volatility}_{classic} \quad - \quad \text{volatility}_{sort}$$

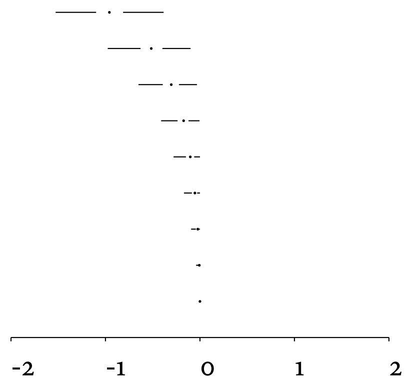
Thus, a negative number means that *classic* is better. Clearly, these numbers will always be negative since in-sample, *sort* cannot be better than *classic*. But more relevant is the actual difference, for which we evaluate the chosen portfolio at the true variance–covariance matrix:

$$\text{actual difference} \quad = \quad \text{actual volatility}_{classic} \quad - \quad \text{actual volatility}_{sort}$$

The hoped-for advantages are shown in the following table. The results are quite intuitive: when assets are strongly correlated, then the advantage of computing a minimum-variance portfolio should be small. But if assets are weakly correlated, we should expect more gains, and indeed, with a low correlation of 0.1, we expect to reduce portfolio risk by one percentage point more than through *sort*.

These tables were produced with the function `qTable` from the `NMOF` package (Gilli et al., 2011).

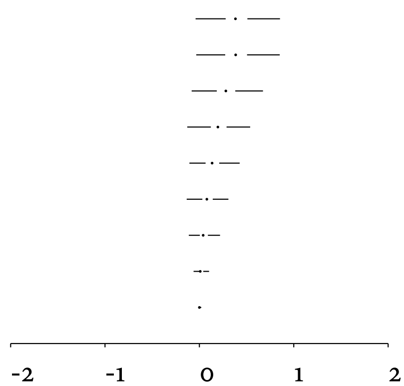
ρ	median	Q_5	Q_{95}
0.1	-1.0	-1.3	-0.6
0.2	-0.5	-0.8	-0.3
0.3	-0.3	-0.5	-0.1
0.4	-0.2	-0.3	-0.1
0.5	-0.1	-0.2	-0.0
0.6	-0.1	-0.1	-0.0
0.7	-0.0	-0.1	-0.0
0.8	-0.0	-0.0	0.0
0.9	-0.0	-0.0	0.0



Distribution of differences between expected (in-sample) volatility: negative values indicate that *classic* promises lower risk than *sort*. By construction, all differences must be negative.

But when we look at the actual differences, these gains disappear. The actual volatility of *sort* is almost never higher than that of *classic*, but is often substantially lower (up to half a percentage point) for low actual correlation. For the empirically more plausible cases of higher correlation, there is essentially no difference.

ρ	median	Q_5	Q_{95}
0.1	0.4	0.1	0.7
0.2	0.4	0.1	0.7
0.3	0.3	0.1	0.6
0.4	0.2	0.0	0.4
0.5	0.1	-0.0	0.3
0.6	0.1	-0.0	0.2
0.7	0.0	-0.0	0.2
0.8	0.0	-0.0	0.1
0.9	0.0	-0.0	0.1



Distribution of differences between actual volatility: negative values indicate that *classic* has lower risk that *sort*.

Discussion

The presented results show that the forecast error for correlation can easily be so large that only relying on marginal risk seems a good prescription for portfolio choice. Thus, Taking-the-Best worked well: selecting the assets that looked best according to a single, easy-to-compute criterion.

This does not imply that we should altogether disregard the comovement of assets. The above example showed just one particular case, an artificial example. We need to empirically analyse whether, where and when the result applies (the driver of the results are variation in the cross-section and forecast error). But the example makes clear that basing the decision how to diversify naïvely on an estimate/forecast of correlation is probably a bad idea. So to stress that point: I do not argue against diversification, but against diversification that is driven by forecasts of correlation.

The result implies, for instance, that simple sorting rules or cutoff rules are likely 'more optimal' than is sometimes thought. At the least, such rules can be used as benchmarks or as the basis of passively-managed portfolios. For active managers, the implication is that finding the right assets should be key (and not fine-tuning asset weights).

The argument was made for equities in a long-only setting, but how about long–short portfolios? It is less clear how a sorting rule might include short positions. At least when absolute position sizes are constrained, allowing short positions in a portfolio may help in reducing risk. But we should keep in mind that investors do not like low risk *per se*. The long-only minimum-variance portfolio is popular in the literature because several empirical studies have shown that it rewards its holder with reasonable returns. There is less empirical evidence for how long–short minimum-variance portfolios behave. (And even less evidence for the *sort* rule!)

Finally, it should be stressed that the underlying idea of this note is not new. A large number of studies in various disciplines, from econometrics to psychology, document the fact that simple methods (as opposed to 'sophisticated' ones) work well for prediction under uncertainty. Sadly, within their respective disciplines, these studies rarely represent mainstream thinking.

References

Edwin J. Elton and Martin J. Gruber. Risk reduction and portfolio size: An analytical solution. *Journal of Business*, 50(4):415–437, 1977.

Gerd Gigerenzer and Daniel Goldstein. Betting on one good reason: The Take The Best heuristic. In Gerd Gigerenzer, Peter M. Todd, and The ABC Research Group, editors, *Simple heuristics that make us smart*. Oxford University Press, 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.

Manfred Gilli, Dietmar Maringer, and Enrico Schumann. *Numerical Methods and Optimization in Finance*. Academic Press, 2011.

Friedrich Leisch. Sweave: Dynamic generation of statistical reports using literate data analysis. In Wolfgang Härdle and Bernd Rönz, editors, *Compstat 2002 – Proceedings in Computational Statistics*, pages 575–580. Physica Verlag, 2002.

Richard O. Michaud. The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

Berwin A. Turlach and Andreas Weingessel. *quadprog: Functions to solve Quadratic Programming Problems.*, 2011. R package version 1.5-4 (S original by Berwin A. Turlach. R port by Andreas Weingessel <Andreas.Weingessel@ci.tuwien.ac.at>.).

Notes

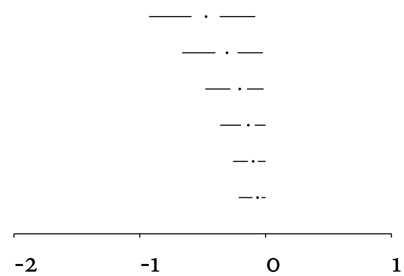
6 Feb 2013

Manfred Gilli suggested to test a more realistic data-generating process in which correlations are not pairwise constant. So I ran a number of experiments. The setup remains as described before, but correlations are now drawn randomly from a range, such as 0.1 to 0.4.†

The following tables show that the overall result remains the same; on average the realised volatility of *sort* is still lower than that of *classic*. But the outcomes are more variable.

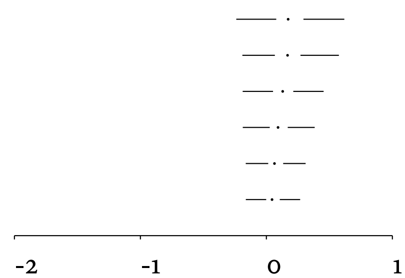
†Since such random matrices are typically not positive-semidefinite, I repair them by replacing their negative eigenvalues as described in Gilli et al. (2011, pp. 396–402).

ρ	median	Q_5	Q_{95}
0.1–0.4	-0.5	-0.8	-0.2
0.2–0.5	-0.3	-0.6	-0.1
0.3–0.6	-0.2	-0.4	-0.1
0.4–0.7	-0.1	-0.3	-0.0
0.5–0.8	-0.1	-0.2	-0.0
0.6–0.9	-0.1	-0.2	-0.0



Distribution of differences between expected (in-sample) volatility: negative values indicate that *classic* promises lower risk than *sort*. By construction, all differences must be negative.

ρ	median	Q_5	Q_{95}
0.1–0.4	0.2	-0.1	0.5
0.2–0.5	0.2	-0.1	0.4
0.3–0.6	0.1	-0.0	0.4
0.4–0.7	0.1	-0.1	0.3
0.5–0.8	0.1	-0.1	0.3
0.6–0.9	0.0	-0.1	0.2



Distribution of differences between actual volatility: negative values indicate that *classic* has lower risk that *sort*.

The most realisitic case is probably the last line in the table; there seems little difference between *classic* and *sort*.

17 Feb 2013

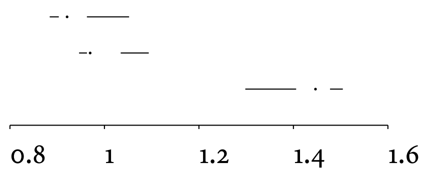
Let me be the first to present empirical evidence against the *sort* rule (but you will see shortly why I find little trouble in reporting this evidence).

Dataset are the components of the HDAX (110 series) and 20 series from the EUROSTOXX 50 universe; altogether 130 series of daily prices. The out-of-sample series start in January 2011 and end in February 2013. I ran 1000 walk-forwards with the following settings: for a given walk-forward, a subset of between 50 and 130 series is chosen randomly. Along this walk-forward, every backward looking window is set randomly to between 180 and 260 trading days; the holding period is set randomly to between 10 and 60 trading days. The cardinality of *sort* is the smallest integer not smaller than

$$\max\left(1/5\{\text{number of assets}\}, 20\right).$$

The next table presents the out-of-sample daily risk (not differences) in percentage points; eg, '1.1' means 1.1% daily volatility.

ρ	median	Q_5	Q_{95}
classic	0.9	0.9	1.0
sort	1.0	1.0	1.1
random	1.4	1.3	1.5



Distributions of actual daily volatility in percentage points.

So for this dataset and this time period *classic* comes out ahead of *sort*. The average advantage when doing a paired comparison is about 4bp per day, which sums to more than half a percentage point when annualised. That is a meaningful reduction in risk, though *sort*'s results are still impressive, given how much simpler than *classic* it is. To give some concrete numbers, a typical outcome might be an annualised 15.2% for *classic*, 15.8% for *sort* and 23.6% for a randomly-chosen equal-weight portfolio with the same cardinality as *sort*. For comparison, the HDAX had an annualised volatility of 23.5% over this period; the EUROSTOXX 50 had about 24.8%.