COMISEF
Computational Optimization Methods in Statistics, Econometrics and Finance

- Marie Curie Research and Training Network funded by the EU Commission through MRTN-CT-2006-034270 -

COMISEF WORKING PAPERS SERIES

WPS-010 15/06/2009

Optimal enough?

M. Gilli
E. Schumann

Optimal enough? ⋆

Manfred Gilli [a] and Enrico Schumann [b,*]

[a] *Department of Econometrics, University of Geneva and Swiss Finance Institute*
[b] *Department of Econometrics, University of Geneva*

Abstract

An alleged weakness of heuristic optimisation methods is the stochastic character of their solutions. That is, instead of finding a truly optimal solution, they only provide a stochastic approximation of this optimum. In this paper we look into a particular application, portfolio optimisation. We demonstrate two points: firstly, the randomness of the 'optimal' solution obtained from the algorithm can be made so small that for all practical purposes it can be neglected. Secondly, and more importantly, we show that the remaining randomness is swamped by the uncertainty coming from the data. In particular, we show that as a result of the bad conditioning of the problem, minor changes in the solution lead to economically meaningful changes in the solution's out-of-sample performance. The relationship between in-sample fit and out-of-sample performance is not monotonous, but still, we observe that up to a point better solutions in-sample lead to better solutions out-of-sample. Beyond this point, however, there is practically no more cause for improving the solution any further, since any improvement will only lead to unpredictable changes (noise) out-of-sample.

Last changes: 14 June 2009

Key words: Optimisation heuristics, Portfolio Optimisation, Threshold Accepting

JEL codes: C61, G11

⋆ Both authors gratefully acknowledge financial support from the EU Commission through MRTN-CT-2006-034270 COMISEF; the data set was provided by DynaGest SA.

* Corresponding author: Department of Econometrics, University of Geneva, Bd du Pont d'Arve 40, 1211 Geneva 4, Switzerland. Tel.: +41 22 379 8218; fax: +41 22 379 8299.

Email addresses: Manfred.Gilli@unige.ch (Manfred Gilli), Enrico.Schumann@unige.ch (Enrico Schumann).

1 Introduction

The aim of portfolio selection is to derive decision rules that help investors to allocate their wealth. The best known of these rules, mean–variance selection (Markowitz, 1952), is often criticised for failing to take into account the non-Gaussian nature of financial time series. Hence alternative and – at least theoretically – superior models have been proposed, with selection criteria that take into account 'fat tails' or asymmetric return distributions. Unfortunately, many of these alternative models lead to optimisation problems that are much harder to solve, and are often infeasible for standard optimisation techniques like linear or quadratic programming. For several specific models, like Expected Shortfall (Rockafellar and Uryasev, 2000; Bertsimas et al., 2004), exact solutions are available, but unfortunately, these techniques are in general not flexible and often cannot accommodate even minor changes in the model; neither can they generally handle realistic constraints like cardinality restrictions or limits to transaction costs.

There is an alternative approach to solve such optimisation problems – heuristics. The term 'heuristic' is used in different scientific areas, with different but related meanings. Mathematicians use it to describe an explanation that is actually not correct (at least not provable), but leads to a correct conclusion nonetheless; in the language of psychologists, heuristics are 'rules of thumb' for decision making that, though sometimes seemingly crude, work robustly in many circumstances (Gigerenzer, 2004, 2008). Winker and Maringer (2007) (following Barr et al. (1995) and similarly Zanakis and Evans (1981)) characterise the term 'heuristic optimisation' through several criteria:

- The method should produce 'good' stochastic approximations of the true optimum, where 'good' is measured in terms of solution quality and computing time.
- The method should be robust in case of comparatively small changes to the given problem, and also for changes in the parameter settings of the heuristic itself. Robustness, again, is measured in (changes in) solution quality and computing time.
- The technique should be easy to implement.
- Implementation and application of the technique should not require subjective elements.

For many techniques like Genetic Algorithms (Holland, 1992) or Simulated Annealing (Kirkpatrick et al., 1983) a considerable theoretical background is available, including mathematical analysis of their convergence. More importantly, heuristics have been shown to work well for problems that are completely infeasible for classical optimisation approaches (Michalewicz and Fogel, 2004).

Heuristics are, with only few exceptions (like Tabu Search), stochastic algorithms. Thus repeated runs, called restarts, of the optimisation algorithm will result in different solutions – even for identical starting points. It is this stochastic nature of the solutions that sometimes causes distrust, for how can we judge whether we have actually found a good solution?

When it comes to the stochastics of solutions, there are similarities and differences between heuristics and classical, deterministic methods. For a given problem with multiple local minima – and very few practical problems are truly convex – repeated runs *with different starting points* will result in different solutions, for both types of techniques. If we characterise a solution (in our case, a portfolio) by its associated objective function value, we may consider the result of one restart as a realisation of a random variable with some unknown distribution. The shape of this distribution will depend on the chosen method, hence when we try to solve a given multimodal problem with a deterministic method, the probability of obtaining, from one optimisation run, a solution of a given quality remains fixed (though unknown). Heuristics on the other hand can move away from local minima, hence allowing more iterations per restart generally changes the shape of the distribution. If the algorithm works properly, with an increasing number of iterations, the distribution becomes steeper (the solutions become less dispersed), and moves to the left, closer to the global minimum. In other words, with a heuristic, the probability for obtaining a solution of a given quality will also depend on the computational resources (iterations) spent on the problem.

In this paper we solve a portfolio optimisation problem with a heuristic technique called Threshold Accepting (Dueck and Scheuer, 1990; Dueck and Winker, 1992), a descendant of Simulated Annealing. Since our main argument does not relate to a specific technique, we describe the algorithm informally here. Threshold Accepting (TA) builds on a simple concept in optimisation called local search. A local search starts with a random feasible solution (that is, a random portfolio) which we call the 'current solution', representing the best solution we have so far. Then again randomly, a new solution close-by is chosen. 'Close-by' means that we slightly perturb the weights of the portfolio (eg, we 'sell' a small quantity of one asset, and invest the proceeds in another asset). This new solution is called a neighbour. If it is better than the current solution, the new solution is accepted and becomes the current solution, if not, it is rejected, and another neighbour is selected. This procedure stops after a preset number of iterations.

Since a local search stops at the first local minimum encountered, TA makes a small adjustment to the procedure: when the algorithm evaluates a neigh-

bour solution, it may also accept this new solution if it is worse than the current solution – as long as the impairment does not exceed a given threshold (hence the method's name). This threshold is set rather generously initially, so that the algorithm may move freely in the search space. Over time, the thresholds are decreased; hence the algorithm gets more select and finally turns into a local search. For a detailed description of TA, see Winker (2001).

So, how to judge the quality of a solution obtained from a TA run? For some problems, the objective function can directly be interpreted. The simplest approach then is to look at the realised value of the objective function associated with the returned solution. Assume we minimise a portfolio's Value-at-Risk, then we could compare our results with the Value-at-Risk of a benchmark portfolio. There is an important empirical question to be asked: how is the in-sample 'performance' that our solution achieves related to its out-of-sample performance? Much research in portfolio optimisation relates to in-sample properties of different methods: given a data set, we aim to minimise drawdown, or ratios of losses to gains. But what we actually want to minimise is *future* drawdown, or the ratio of *future* losses to *future* gains. There exists comparatively little research into how these objectives relate to the quantities that we actually optimise.

The main point that we will stress in this paper is the following: the aim of portfolio selection is to find decision rules that tell investors how to invest. To this purpose, we set up a model (a selection criterion and constraints), but this model will depend on future asset prices which we do not know. Hence, we need to approximate/estimate/simulate these prices, and then solve the model. Zanakis and Evans (1981, p. 85) list cases where a heuristic should be the method of choice, there first case being '[i]nexact or limited data used to estimate model parameters [that] may inherently contain errors much larger than the "suboptimality" of a good heuristic.' We will argue that portfolio selection belongs to this class of optimisation problems.

We show that a heuristic technique can give a good solution to the model. Better solutions in-sample (ie, solutions of the model) also lead to better solutions out-of-sample (the actual problem). But this holds true only up to a certain point: beyond this point there is no more cause for improving the model's solution any further, for any improvement does not lead to systematic improvements out-of-sample any more. The remainder of this paper is structured as follows: in Section 2 we briefly describe the optimisation problem and our data. Section 3 and Section 4 then discuss both the in-sample and out-of-sample performance of our algorithm. Section 5 concludes.

2 Data and Methodology

There are $n_{\mathcal{A}}$ risky assets available, with current prices collected in a vector p_0. We are endowed with an initial wealth v_0, and wish to select a portfolio $x = [x_1 \; x_2 \; \ldots \; x_{n_{\mathcal{A}}}]'$ of the given assets. We can thus write down a budget constraint

$$v_0 = x' p_0 \,.$$

The vector x stores the number of shares or contracts, that is, integer numbers.

The chosen portfolio is held for one period, from now (time 0) to time T. End-of-period wealth is given by

$$v_T = x' p_T \,,$$

where the vector p_T holds the asset prices at T. Since these prices are not known at the time when the portfolio is formed, v_T will be a random variable, following some unknown distribution. It is often convenient to rescale v_T to a return r_T, that is

$$r_T = \frac{v_T}{v_0} - 1 \,.$$

Let $\mathcal{J}$ be the set of assets in the portfolio, then our optimisation problem can be written as

$$\begin{aligned} &\min_x \; \Phi \,, \\ &x_j^{\text{inf}} \leq x_j \leq x_j^{\text{sup}} \qquad j \in \mathcal{J} \,, \\ &K_{\text{inf}} \leq \#\{\mathcal{J}\} \leq K_{\text{sup}} \,. \end{aligned}$$

We use the downside semi-variance as our objective function Φ, computed as

$$\frac{1}{n_{\mathcal{S}}} \sum_{r<r_d} (r_d - r)^2 \,.$$

Here r are the sample returns of a given portfolio, r_d is a desired-return threshold (which we set to zero), and $n_{\mathcal{S}}$ stands for the number of observations (or scenarios). We do not include minimum-return constraints, thus we solely minimise the risk of the portfolio (defined as the semi-variance). This is equivalent to assuming equal means for all assets. There is strong empirical evidence that such an approach gives good results in out-of-sample tests (see Board and Sutcliffe (1994); Chan et al. (1999) for variance-minimisation, Gilli and Schumann (2009) for alternative risk func-

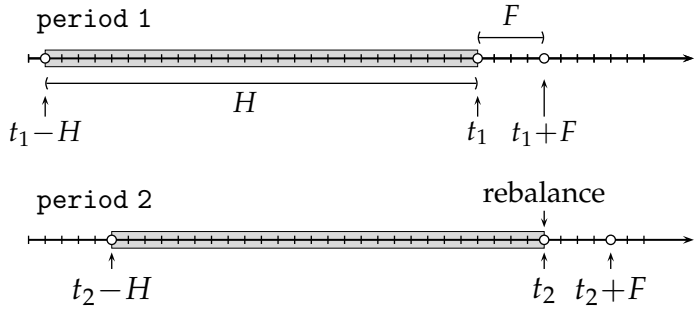
tions).

x_j^{inf} and x_j^{sup} are minimum and maximum holding sizes, respectively, for those assets included in the portfolio (ie, those in $\mathcal{J}$). K_{inf} and K_{sup} are cardinality constraints which set a minimum and maximum number of assets in $\mathcal{J}$. We set $x^{\text{inf}} = 1\%$ and $x^{\text{sup}} = 5\%$, thus we do not allow short positions. An upper cardinality is set to 50;. the value of K_{inf} is set to 10. We do not include a riskless asset. Since our algorithm works with actual position sizes, that is integer numbers, a small fraction of less than 1% of the portfolio is usually left uninvested.

The data set comprises more than 500 price series of European companies from the Dow Jones STOXX universe, at a daily frequency, spanning the period from March 2000 to March 2008. All stocks are denominated in euro.

Moving windows

We conduct rolling-window backtests with an historical window of length H, and an out-of-sample holding period of length F. We set H to one year, F to three months. Thus we optimise at point in time t_1 on data from $t_1 - H$ to $t_1 - 1$, the resulting portfolio is held until $t_2 = t_1 + F$. At this point, a new optimal portfolio is computed, using data from $t_2 - H$ until $t_2 - 1$, and the existing portfolio is rebalanced. This new portfolio is then held until $t_3 = t_2 + F$, and so on. This is illustrated in the following figure.



With our data set, we have 35 investment periods, that is we optimise the first time on 27 March 2000 (t_1), the last date is 2 January 2008 (t_{35}). From the historical data in every period (the in-sample data) we create scenarios via the bootstrapping procedure described in Gilli and Schumann (2009). The created scenarios are saved for every period so that all optimisation runs are conducted on the same data. The out-of-sample data for every period is the actual, historical data.

3 In-sample results

The portfolio optimisation is handled with Threshold Accepting (Winker, 2001; Gilli et al., 2006). We measure the computational resources that the optimisation algorithm employs by the total number of iterations (ie, the total number of objective function evaluations). With an increasing number of iterations the average quality of a solution per restart should improve while the solutions' variability should decrease; eventually the solutions' distribution should collapse into a single point (the global minimum) as the number of iterations goes to infinity. In other words, if we run a TA with I_1 iterations, and one with I_2 iterations, where $I_1 < I_2$, then on average the TA with I_2 iterations will give better solutions. In-sample, thus, we face a trade-off between solution quality and computational resources spent.

We test the performance of TA with 1, 1 000, 5 000, 15 000, 50 000 and 100 000 iterations. With only 1 iteration, we actually obtain random (but feasible) portfolios. Hence we will often write 'random portfolios' instead of 1 iteration. Setting the computational resources fixes the distributions from which we draw our solutions. To approximate these distributions, we run the algorithm 100 times for every level of iterations, and compute the empirical cumulative distributions functions as estimates of the true distributions. Figure 1 shows the empirical results for the first four periods, for later periods we obtain similar results.

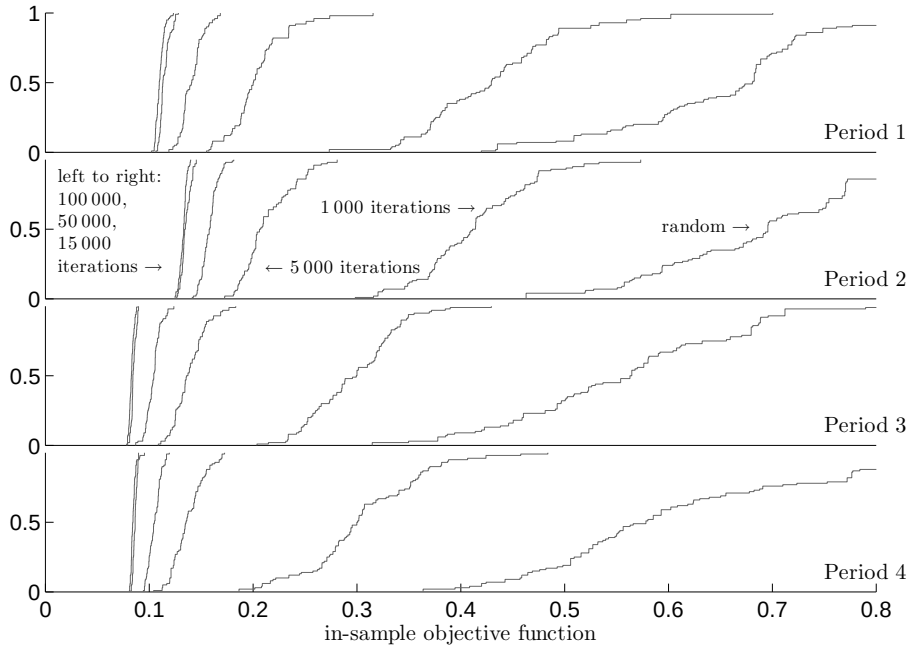


Figure 1. In-sample convergence.

With an increasing number of iterations the distributions become rapidly steeper and move to the left. There is virtually no difference any more

between the results for 50 000 and 100 000 iterations. Thus we can make the randomness of the in-sample objective function very small. To give a concrete example: in period 4 (the lower panel in Figure 1) the objective function ranges between 0.08 and 0.09 for 100 000 steps, while for random portfolios there is a range between 0.36 and 0.90. The objective function is the semi-variance of the portfolio, so taking the square root, 0.08 to 0.09 translates into a daily downside-deviation of between 0.28% and 0.30%. Of course, we have no guarantee that we really have found the global minimum. But with a bound at zero, and the best random portfolios having about double the variation of our optimised portfolios (ie, a four times higher objective function), TA seems to consistently find 'good' solutions. In other words, TA seems well capable of solving our model.

4 Out-of-sample results

A given portfolio maps not only into an in-sample objective function, but also into an out-of-sample return. This link between in-sample fit and actual out-of-sample performance is very noisy, though. Figure 2 shows a scatter plot of in-sample objective functions against out-of-sample returns for one investment period. (We will only look at out-of-sample returns here. The exercise could be easily extended to include other performance measures.)

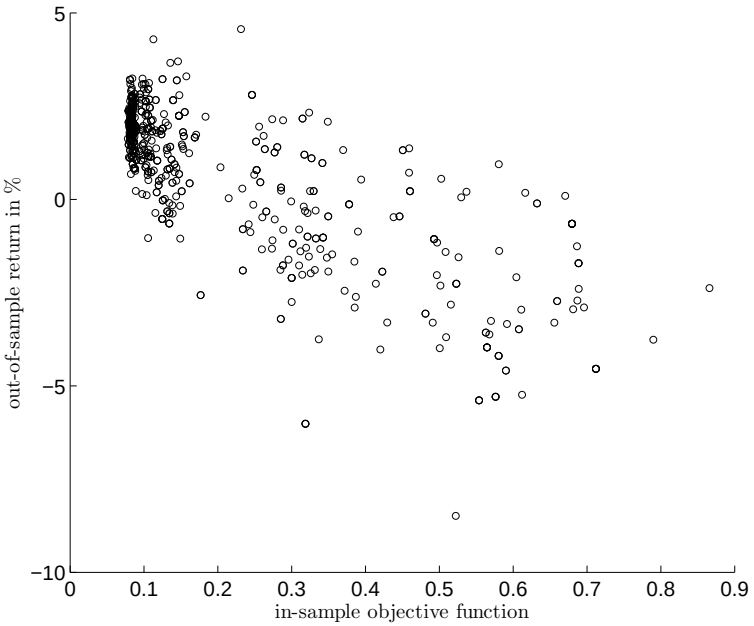


Figure 2. In-sample objective functions vs out-of-sample return for one period.

The picture shows several characteristic features that we also found in the other periods: for high in-sample objective functions (associated with the

random portfolios), the returns are widely scattered. When we move to the left, closer to the in-sample optimum, the cloud becomes denser. There is no clear monotonous relationship between in-sample fit and out-of-sample return. In particular, the leftmost part of the cloud, which covers only a narrow range in terms of the in-sample objective function, still gives rise to highly variable out-of-sample returns. Thus, even though the in-sample stochasticity is small, the out-of-sample randomness is disproportionately larger, which shows the high sensitivity of the model with respect to small changes in the solution.

To be more concrete: the objective function values of the 10 best solutions for the pictured period cover a range from 0.0783 to 0.0797, which means portfolios with a daily downside-deviation between 0.2779% and 0.2824%. (We give 4 digits to illustrate how small the differences are, not because we think that working with such 'precision' is a good idea.) These minuscule differences translate into out-of-sample returns between 1.13% and 2.37% over the following 3-month period. In other words, less than half a basis point in-sample leads to a difference of more than one percentage point out-of-sample.

To better understand the link between in-sample and out-of-sample, we compute the performance of 'rank-portfolios'. Rank-portfolios are constructed as follows. In every period, with iterations set to six different levels (1, 1 000, 5 000, 15 000, 50 000 and 100 000), and 100 restarts for each level, we have a total of 600 portfolios. We sort these portfolios according to their in-sample objective function. A natural decision rule is to select the best in-sample portfolio, that is the portfolio with the lowest objective function. We call this the rank-1 portfolio. Likewise, we can determine portfolios ranked from 2 to 600 for every period. This is illustrated in Figure 3, where we have indicated the rank-1 and the rank-200 portfolios for the periods one to four.

While the rank is determined in-sample, let us write r_t^i for the out-of-sample performance of the portfolio with rank i in period t. The total out-of-sample return R_T^i along the path of the ith rank portfolio is

$$R_T^i = \prod_{t=1}^{T} 1 + r_t^i$$

with T the number of periods. Hence R_T^1 is the total return for an investor who always chooses the best (rank-1) in-sample portfolio, R_T^2 is the total return for an investor who always chooses the second-best (rank-2) in-sample portfolio, and so on, until R_T^{600} gives the total return for the worst in-sample portfolios.

Figure 4 plots the growth of €1 invested in the rank-1 portfolio and the

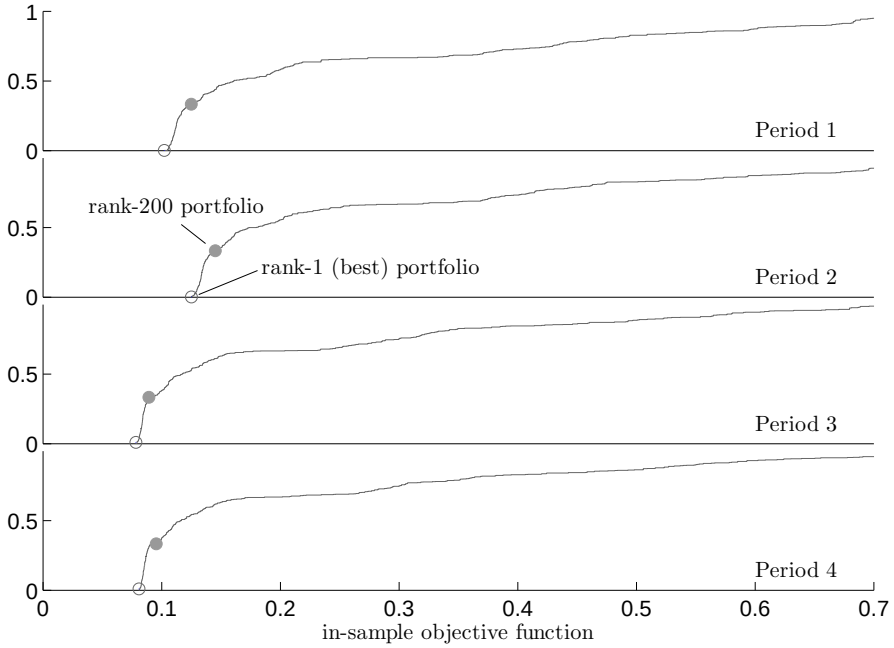


Figure 3. In-sample objective functions for the first four periods.

growth of € 1 invested in the rank-600 portfolio (ie, the paths of R_t^1 and R_t^{600} for $t = 1, \ldots, T$). The in-sample-best portfolio clearly dominates the in-sample-worst portfolio. The total return difference is more than 6% per year.

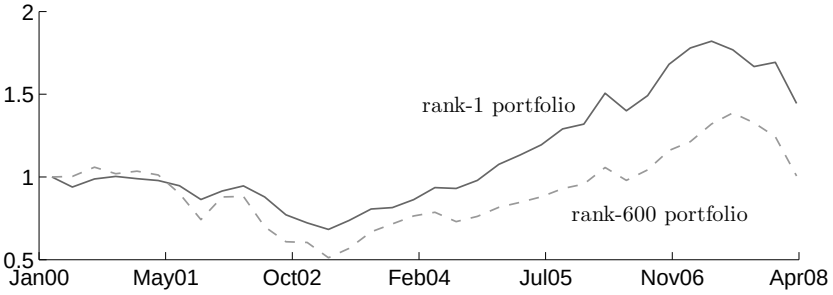


Figure 4. Paths.

Figure 5 plots the total return $R_T^i - 1$ (annualised) for all 600 rank portfolios, plotted against the respective rank. While there is a negative relationship between rank and out-of-sample performance, we can see that this relation is far from monotonous. (Again, we only look at returns. Given that returns range between 0% and 8%, it seems unlikely that any risk-adjustment would change the conclusion.) We regress the out-of-sample return of a portfolio on its in-sample rank; the following table shows the results. (A linear regression is certainly an inappropriate tool to measure the actual relationship between rank and return, but it underlines the looseness of the relation as becomes manifest in the small estimated slopes, and

large residuals.)

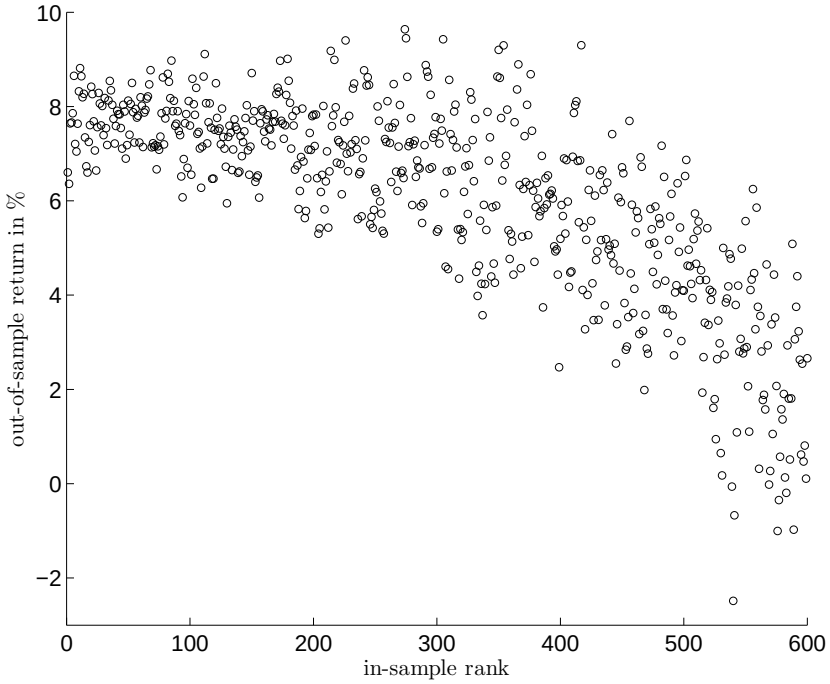


Figure 5. In-sample rank (lower is better) vs out-of-sample return.

ranks	constant in %	slope in %	std. of residuals in %
1–100	7.72	-0.0005	0.60
101–200	7.55	-0.0025	0.72
201–300	6.98	0.0023	1.06
301–400	6.75	-0.0092	1.47
401–500	5.96	-0.0168	1.37
501–600	4.45	-0.0311	1.70

We see that for the portfolios with ranks 1 to 200 the slope is one-fourth of a basis point or less. So advancing 100 ranks would, on average, result in an improvement of 0.25% in total return. For the best 100 ranks, this improvement would be less than 0.05%. Given the enormous uncertainty in the data, reflected in a standard deviation of the residuals of 0.60%, this 'improvement' can economically be regarded as zero.

We can also conduct a more direct test: assume there are two investors, A and B, both use TA to find portfolios. Investor A always lets his algorithm run for 100 000 iterations, while B uses only one iteration (ie, random portfolios). Going back to Figure 3, A will always pick portfolio from the left-

most distribution, while B will choose from the more dispersed random portfolios. In-sample, A's portfolios will thus look much better than B's, but how likely is A to outperform B? Given the noisy link in-sample to out-of-sample, we will unlikely observe that A's portfolio dominates B's in every period (not even to a higher order), but maybe it 'almost dominates' (Leshno and Levy, 2002).

We start by randomly drawing one period from the $T = 35$ periods available. As explained before, each period can be split into an in-sample and an out-of-sample part. Then, for A, we randomly pick one portfolio out of the 100 that were optimised with 100 000 iterations on the in-sample part of this period, and record its out-of-sample performance r_1^A. We do the same for B, and hence obtain r_1^B. Next we pick again randomly one period (could even be the same period), choose again a portfolio for A and one for B, and record their out-of-sample returns, r_2^A and r_2^B. After p such draws, A's normalised net worth will be

$$(1 + r_1^A)(1 + r_2^A)(1 + r_3^A) \dots (1 + r_p^A)\,,$$

and B's will have grown to

$$(1 + r_1^B)(1 + r_2^B)(1 + r_3^B) \dots (1 + r_p^B)\,.$$

We then compute the geometric outperformance of A as

$$g = \frac{\prod_{i=1}^{p} 1 + r_i^A}{\prod_{i=1}^{p} 1 + r_i^B} - 1\,.$$

We repeat this exercise 10 000 times, for different levels of p. Thus, for a fixed investment horizon p, we obtain a distribution of g and can now compare how much better A's portfolio performed compared with B's.

Figure 6 shows these distributions of g. A value of $g = 1$ means that A's wealth has grown to twice the wealth of B, while $g = 0$ indicates that A and B are equally wealthy. One period ($p = 1$) in our setting was three months, hence 5 years are 20 periods ($p = 20$), and 20 years are 80 periods ($p = 80$). There is only a probability of 57% that A outperforms B after one period (3 months) (ie, a probability that $g > 0$). After 5 years, the probability is 78%, and even after 20 years it is 'only' 94%. Note, however, that these distributions are not symmetric.

Now let B switch to a TA with 50 000 iterations, half the computational resources that A employs. Figure 6 gives a clear answer: even after 20 years, the distribution of g is symmetric around $g = 0$, hence there is little predictable difference between the portfolios of A and B.

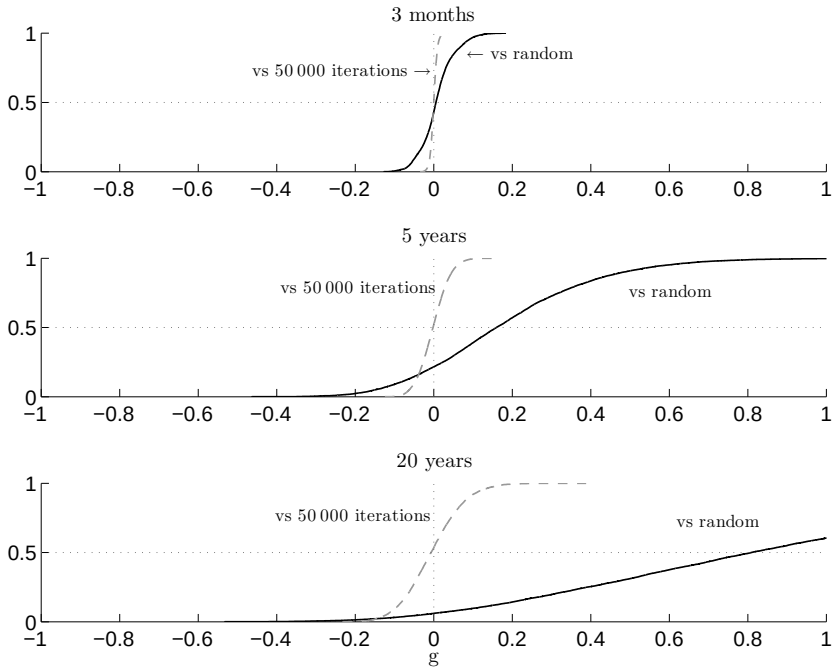


Figure 6. Distribution of outperformance of portfolios with 100 000 iterations.

5 Conclusion

In this paper we analysed the stochastics of solutions obtained from a heuristic optimisation method when applied to a financial problem, portfolio optimisation. Our findings indicate that the randomness added as a result of using a heuristic can be made so small that it practically becomes irrelevant. The true uncertainty stems from the sensitivity of the model: even small changes in a portfolio's composition, reflected in minuscule differences in the in-sample objective function, can lead to completely disproportionate differences in the out-of-sample performance. Any 'exact' method would suffer from the same problem, while giving the false impression of having provided the 'optimal' solution.

The advantage of heuristics here is that they give the analyst more freedom when setting up the optimisation model in the first place, since heuristics can accommodate alternative ways to model the data, or alternative objective functions. Thus, heuristics may be a valuable tool for truly improving the solution not of the model, but of the actual problem.

References

Richard S. Barr, Bruce L. Golden, James P. Kelly, Mauricio G. C. Resende, and William R. Stewart. Designing and Reporting on Computational Experiments with Heuristic Methods. *Journal of Heuristics*, 1(1):9–32, 1995.

Dimitris Bertsimas, Geoffrey J. Lauprete, and Alexander Samarov. Shortfall as a risk measure: properties, optimization and applications. *Journal of Economic Dynamics and Control*, 28(7):1353–1381, 2004.

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.

Gunter Dueck and Tobias Scheuer. Threshold Accepting. A General Purpose Optimization Algorithm Superior to Simulated Annealing. *Journal of Computational Physics*, 90(1):161–175, 1990.

Gunter Dueck and Peter Winker. New concepts and algorithms for portfolio choice. *Applied Stochastic Models and Data Analysis*, 8(3):159–178, 1992.

Gerd Gigerenzer. Fast and Frugal Heuristics: The Tools of Bounded Rationality. In Derek J. Koehler and Nigel Harvey, editors, *Blackwell Handbook of Judgment and Decision Making*, chapter 4. Blackwell Publishing, 2004.

Gerd Gigerenzer. Why Heuristics Work. *Perspectives on Psychological Science*, 3(1):20–29, 2008.

Manfred Gilli and Enrico Schumann. An Empirical Analysis of Alternative Portfolio Selection Criteria. *Swiss Finance Institute Research Paper No. 09-06*, 2009.

Manfred Gilli, Evis Këllezi, and Hilda Hysi. A data-driven optimization heuristic for downside risk minimization. *Journal of Risk*, 8(3):1–18, 2006.

John H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, 1992.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

Moshe Leshno and Haim Levy. Preferred by "All" and Preferred by "Most" Decision Makers: Almost Stochastic Dominance. *Management Science*, 48 (8):1074–1085, 2002.

Harry M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

Zbigniew Michalewicz and David B. Fogel. *How to Solve it: Modern Heuristics*. Springer, 2004.

R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2(3):21–41, 2000.

Peter Winker. *Optimization Heuristics in Econometrics: Applications of Thresh-*

old Accepting. Wiley, 2001.

Peter Winker and Dietmar Maringer. The Threshold Accepting Optimisation Algorithm in Economics and Statistics. In Erricos John Kontoghiorghes and Cristian Gatu, editors, *Optimisation, Econometric and Financial Analysis*, volume 9 of *Advances in Computational Management Science*. Springer, 2007.

Stelios H. Zanakis and James R. Evans. Heuristic "optimization": why, when, and how to use it. *Interfaces*, 11(5):84–91, 1981.