

Performance Measurement via Random Portfolios

Patrick Burns*

2nd December 2004

Abstract

Problems with performance measurement using information ratios relative to a benchmark are exposed. Random portfolios (that obey constraints but disregard utility) are shown to measure investment skill effectively. Investment mandates can also be based on random portfolios—this allows active fund managers more freedom to implement their ideas, and provides the investor more flexibility to gain utility. The issue of the proper attitude towards tracking error is broached, but left largely undecided. There is also a critique of Fisher’s method of combining p-values that shows Stouffer’s method to be preferable.

1 Introduction

The accurate assessment of the skill of fund managers is quite obviously of great value. It is also well known to be a very difficult task. A variety of techniques, some quite clever, have been devised. Some methods measure individual managers, others a class of managers. A few references are [Kosowski et al., 2001], [Muralidhar, 2001], [Engstrom, 2004], [Ding and Wermers, 2004]. There are also [Ferson and Khang, 2002] and [Grinblatt and Titman, 1993].

More accurate performance measurement allows a quicker determination of whether a fund manager has skill or not. It can also provide a more fair method of compensating fund managers for their contribution to the investor.

This paper focuses on using random portfolios [Dawson and Young, 2003] to measure the skill of fund managers, and to specify mandates. Conceptually we want to look at all portfolios that satisfy the constraints, and compare their realized utility to the realized utility of the fund under question. For practical reasons we take a random sample from the set of portfolios satisfying the constraints to use in the comparisons. This leaves us free to use whatever measure (or measures) of quality that we like, and we will have a statistical statement of the significance of the quality of the fund. This procedure eliminates some

*This report can be found in the working papers section of the Burns Statistics website <http://www.burns-stat.com/>. The author thanks Craig Israelsen for providing the data in Table 1; and Larry Siegel and Barton Waring for a helpful discussion.

of the noise that results from assessing a fund's outperformance relative to a benchmark.

R [R Development Core Team, 2004] was used for computations and graphs for this paper. Random portfolios and optimizations were done with the POP Portfolio Construction Suite [Burns Statistics, 2004].

The remainder of the paper consists of:

- Section 2 on measuring funds relative to a benchmark, and some of the problems that ensue.
- Section 3 on combining p-values and using random portfolios to measure fund manager skill.
- Section 4 on some issues with mandates, including a look at tracking error.
- Section 5 on creating mandates based on random portfolios.
- Section 6 that summarizes.

2 Management Against a Benchmark

Currently a great amount of performance analysis is relative to a benchmark. Sometimes this is done because it is deemed reasonable, but other times for lack of an alternative. A good discussion of the use and abuse of benchmarks is [Siegel, 2003].

In this section (and the next) we use a dataset of the daily returns of an unsystematic collection of 191 large-cap and small-cap US equities. The data start at the beginning of 1996 and end after the third quarter of 2004. Results are reported for each quarter except the first two. The first two quarters are excluded so that all results are out of sample.

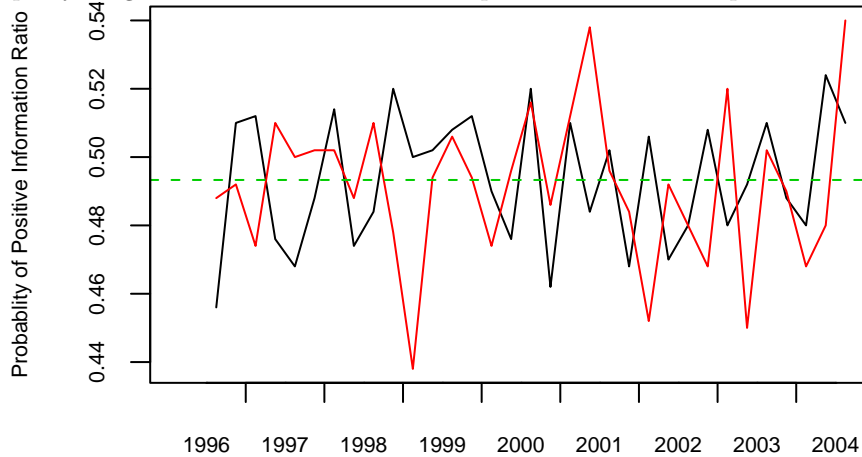
One thousand random portfolios were created from this universe with the constraints that no more than 100 names were in a portfolio, no short values were allowed, the maximum weight of any asset was 10%, and the sum of the 8 largest weights was no more than 40%. In some figures the first 500 random portfolios are compared to the second 500 in order to indicate the significance of any pattern that might appear.

Three artificial benchmarks were created. The first is the equal weighting of the assets. The other two have weights that were randomly generated. These latter two are referred to as the “random benchmarks”—note that the randomness is only in the selection of the weights of the assets, and these weights are held fixed throughout time.

2.1 Outperforming the Benchmark

The information ratios of the random portfolios were calculated relative to the benchmark that has equal weight in each stock. (An information ratio is the

Figure 1: Probability of a positive information ratio by quarter relative to the equally weighted benchmark. Each line represents 500 random portfolios.



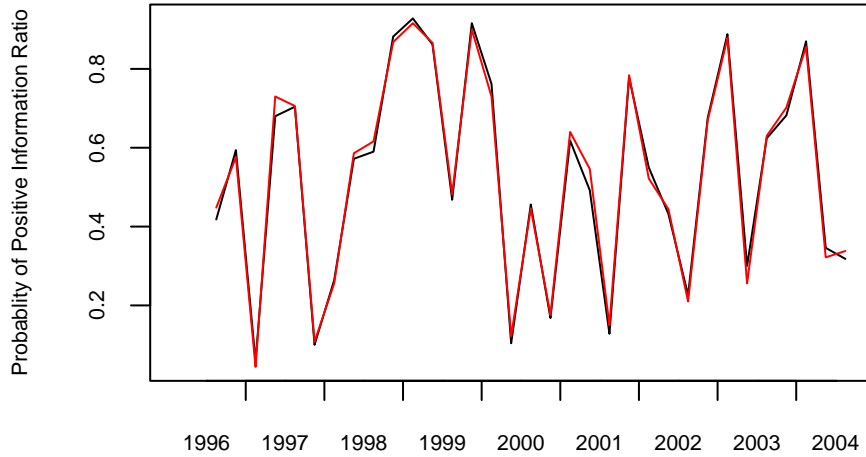
annualized return in excess to the benchmark divided by the annualized standard deviation of the differences in returns—an excess return derived from a regression rather than subtraction is more desirable (see [Siegel, 2003]) but for simplicity is not done here.) Figure 1 shows the probability that the random portfolios have a positive information ratio against this benchmark for each quarter. The black line corresponds to the first 500 random portfolios and the red line to the second 500.

We might have expected the fraction of portfolios that outperform the equally weighted benchmark to be closer to 50%. (The average probability is indicated by the horizontal line.) The p-value is 0.006 for the test that positive and negative information ratios are equally likely. Note though that the benchmark is outside the constraints that we have put on the random portfolios—the portfolios can have at most 100 constituents while the benchmark has 191. The benchmark is likely to have smaller volatility and hence a slight advantage in outperformance.

While there is a tendency for the equally weighted benchmark to outperform, there seems to be no systematic difference between quarters.

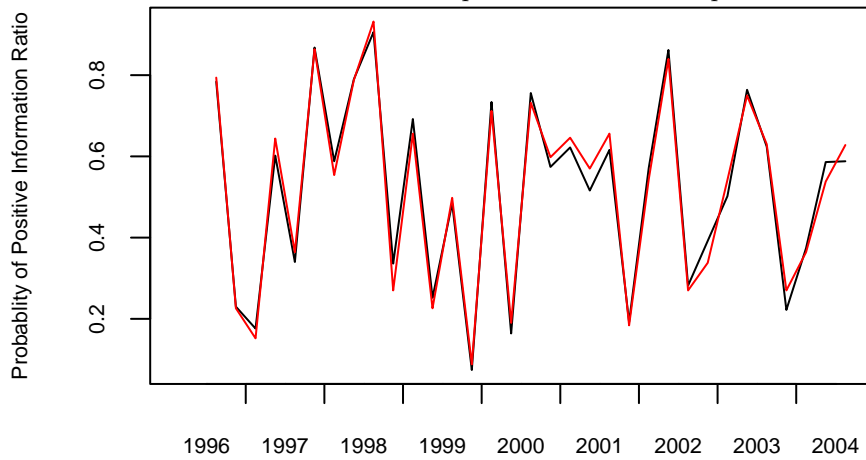
We now look at two benchmarks with randomly generated weights. The mean weight is about 0.5%, and the maximum weight in each benchmark is slightly over 2.5%. Figures 2 and 3 show the probability of a positive information ratio. In these plots there are undeniable differences between quarters. In some quarters there is a strong tendency for the benchmark to outperform the random portfolios, in others a strong tendency for the benchmark to underperform.

Figure 2: Probability of a positive information ratio by quarter relative to the first random benchmark. Each line represents 500 random portfolios.



There is no consistency of outperformance between the two benchmarks.

Figure 3: Probability of a positive information ratio by quarter relative to the second random benchmark. Each line represents 500 random portfolios.



On reflection this result should not be so surprising—though the extent of the effect may be. A benchmark will be hard to beat during periods when the most heavily weighted assets in the benchmark happen to do well. Likewise, when the assets with large weights in the benchmark do relatively poorly, then the benchmark will be easy to beat.

Figure 4 shows the quarterly returns of each of the three benchmarks plotted against each other. The three benchmarks are obviously highly correlated. This seems contradictory since the probabilities of outperforming the benchmarks didn't appear to be related. The explanation is illustrated by Figure 5. This shows the returns of the three benchmarks and the probability of outperformance for each quarter. Even slight differences in return between the benchmarks cause dramatic differences in the probability of outperformance. That is, random portfolios provide a very sensitive measure of performance.

Clearly the more unequal the weights in a benchmark, the more extreme the swings will be in the probability of outperforming. In this regard, the random benchmarks that are used here are not at all extreme compared to many indices that are used in practice as benchmarks.

Table 1 shows a history of U.S. mutual fund outperformance relative to the “best fitting” benchmark of each fund. The data in this table were computed by Craig Israelsen using the Morningstar database. [Israelsen, 2003] alludes to the method of choosing the benchmark for each fund.

There are two histories for the S&P 500—one with all of the available funds, and one containing only the funds that were live in all of the years. This was to explore the possibility of survival bias. Not surprisingly, survival bias appears to be minimal.

The pattern of outperformance of the S&P 500 by the funds is quite similar to that for the random benchmarks as exhibited in Figures 2 and 3—some years a large fraction of funds underperform and other years a large fraction outperform.

Interpreting this data in the way that it is often used, we infer that managers were, in general, bad during the 90's, then they suddenly became very good for three years starting in 2000, then returned to being bad in 2003. This is clearly a ridiculous inference, but nonetheless is often done.

The outperformance of funds relative to the other two benchmarks, while not completely stable, is much less variable. The S&P Midcap 400 almost always beats more than half of the funds that track it, while the Russell 2000 is almost always beat by more than half the funds that track it. There are several possibilities:

- The fund managers that track the S&P Midcap are inept, and the fund managers that track the Russell 2000 are quite skillful.
- The S&P Midcap has been hard to beat and the Russell 2000 has been easy to beat.
- The volatilities of the funds are substantially different from the benchmark volatility.

Figure 4: Scatterplots of quarterly returns of the three hypothetical benchmarks.

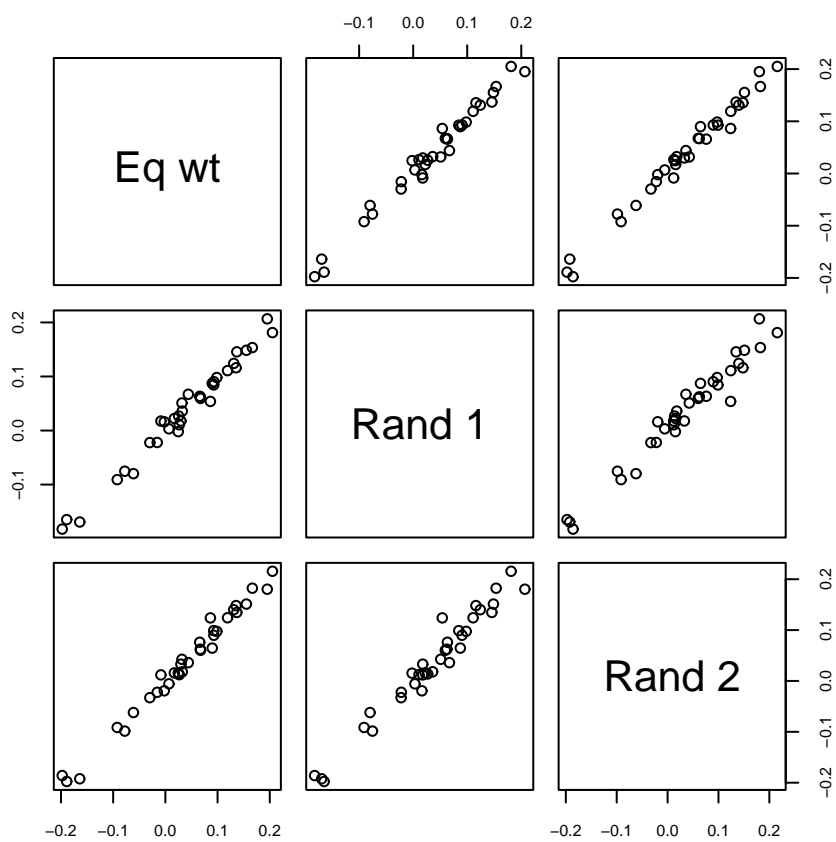
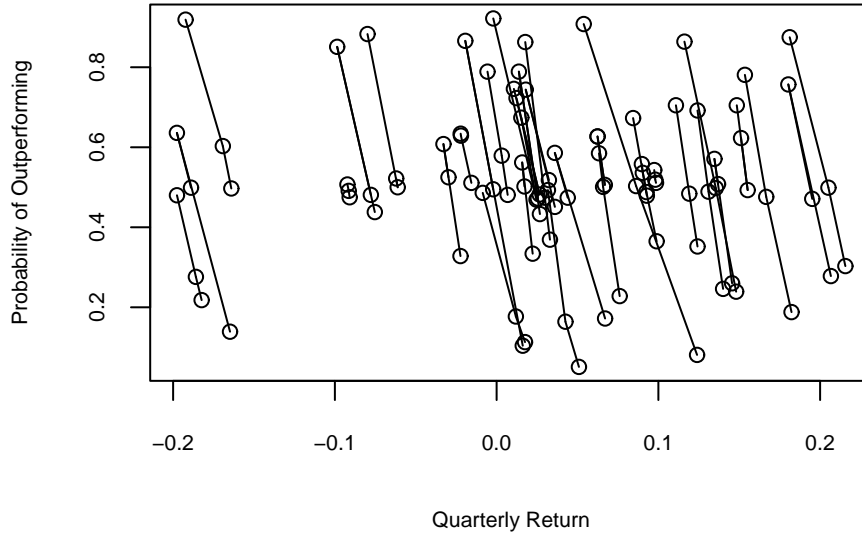


Figure 5: Probability of outperformance by quarterly return for the three hypothetical benchmarks.



- The outperformance is an artifact of the way that benchmarks have been assigned to funds.

We don't have enough information to decide among these. Random portfolios could help inform us.

Some would argue—given the evidence we've just seen—that benchmarks should be equally weighted indices. Even if this were accepted as practical (see [Siegel, 2003] for some reasons why it isn't), it still doesn't solve the issue of accurately measuring skill. Figure 6 shows the probability of the random portfolios having an information ratio greater than two relative to the equally weighted benchmark. There are definite systematic differences by quarter—sometimes a large information ratio is easier to achieve than at other times.

2.2 Information Ratios and Opportunity

Figure 6 implies that the distribution of information ratios changes from quarter to quarter. At least part of the reason is that information ratios are not purely a measure of skill, but rather are a combination of skill and opportunity.

Imagine a case where all of the assets in the universe happen to have the same return over a time period. Portfolios will vary from each other during the period and hence have non-zero tracking error relative to the index. However, all portfolios will end the period with the same return—all information ratios will be zero.

Table 1: US mutual funds outperforming benchmarks. Source: Craig Israelsen

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
S&P 500 return	1.31	37.53	22.94	33.35	28.57	21.04	-9.10	-11.88	-22.09	28.67
# of funds (all)	100	115	126	136	155	173	183	204	212	212
% funds outperform	30	9.6	24.6	14.0	24.5	27.2	72.7	60.8	56.6	13.2
95% conf. int.	(21,40)	(5,16)	(17,33)	(9,21)	(18,32)	(21,34)	(66,79)	(54,68)	(50,63)	(6,20)
# of funds (full history)	100	100	100	100	100	100	100	100	100	100
% funds outperform	30	9	22	13	16	23	74	63	59	12
95% conf. int.	(21,40)	(4,16)	(14,31)	(7,21)	(9,25)	(15,32)	(64,82)	(53,72)	(49,69)	(6,20)
S&P Midcap 400 ret	-3.59	30.92	19.18	32.24	19.11	14.72	17.72	-0.60	-14.53	35.59
# of funds	40	48	52	63	77	95	101	113	121	121
% funds outperform	55.0	41.7	50.0	23.8	27.3	53.7	33.7	39.8	38.0	28.1
95% conf. int.	(38,71)	(28,57)	(36,64)	(14,36)	(18,39)	(43,64)	(25,44)	(31,49)	(29,47)	(20,37)
Russell 2000 return	-1.82	28.44	16.49	22.36	-2.55	21.26	-3.02	2.49	-20.48	47.25
# of funds	26	34	40	50	69	84	97	111	115	115
% funds outperform	61.5	58.8	75.0	56.0	60.9	63.1	71.1	56.8	67.8	41.7
95% conf. int.	(41,80)	(41,75)	(59,87)	(41,70)	(48,72)	(52,73)	(61,80)	(47,66)	(58,76)	(33,51)

Figure 6: Probability of an information ratio greater than two relative to the equally weighted benchmark. Each line represents 500 random portfolios.

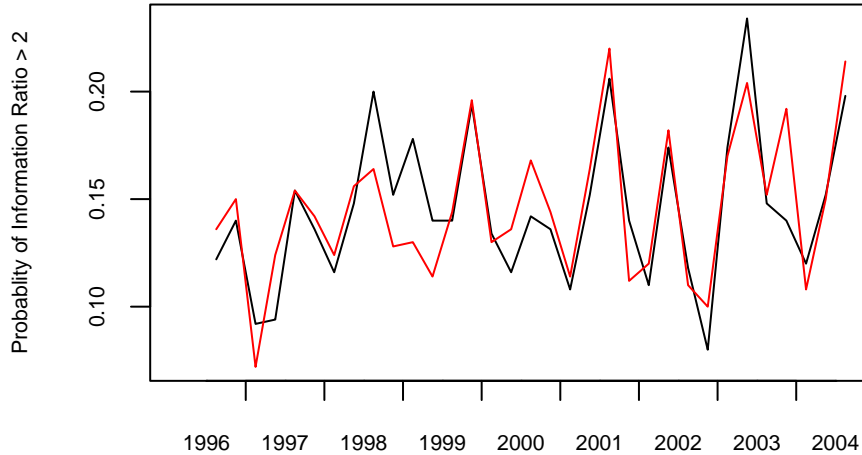


Figure 7 shows the standard deviation of the information ratios of the random portfolios for each quarter and each of the three benchmarks. The naive assumption is that the standard deviations should all be 2. The plot exhibits definite differences between quarters and between benchmarks.

Certainly the cross sectional spread of the full-period returns has an effect on the standard deviation of information ratios. The volatility over time of the assets will also have an effect. Figure 8 shows an experiment of varying these. The data for the first quarter of 2004 were used. Each point in the figure has had the volatility of each asset multiplied by a value and the returns for the period multiplied by a value. The point at (1, 1) corresponds to the real data—there the standard deviation of the information ratios (relative to the equally weighted benchmark) is about 1.8. The points that are at 2 on the horizontal axis have twice the spread of returns as the real data (a stock that really had a 3% return gets a 6% return, and a stock with a -1% return gets a -2% return). The points that are at 0.5 on the vertical axis have half of the volatility as the real data for all of the assets. The point at (2, 0.5) has a standard deviation of information ratios that is about 7.

Figure 8 shows that the cross sectional spread of asset returns is very important to the opportunity to achieve a large information ratio. The spread of returns has a bigger impact as the volatility of the assets decreases. Obviously in reality there is a connection between the volatility of the individual assets and the cross sectional spread of returns, but there is no reason to suppose that they are in lock step.

Figure 7: Standard deviations of the information ratios of random portfolios by quarter.

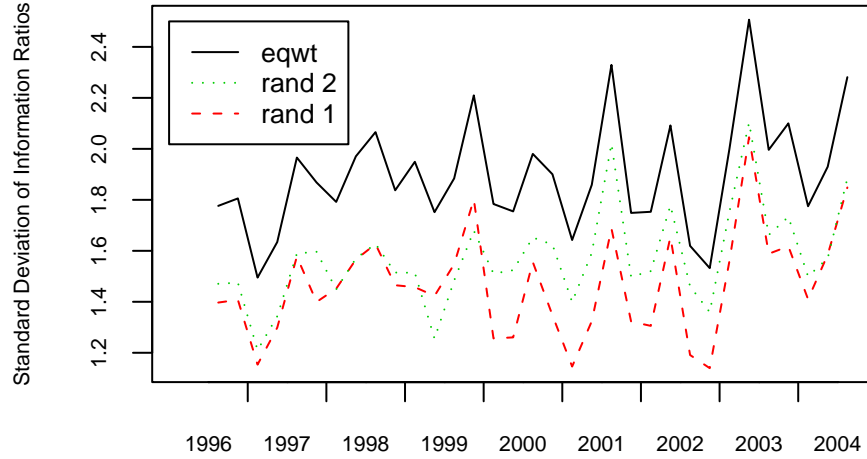
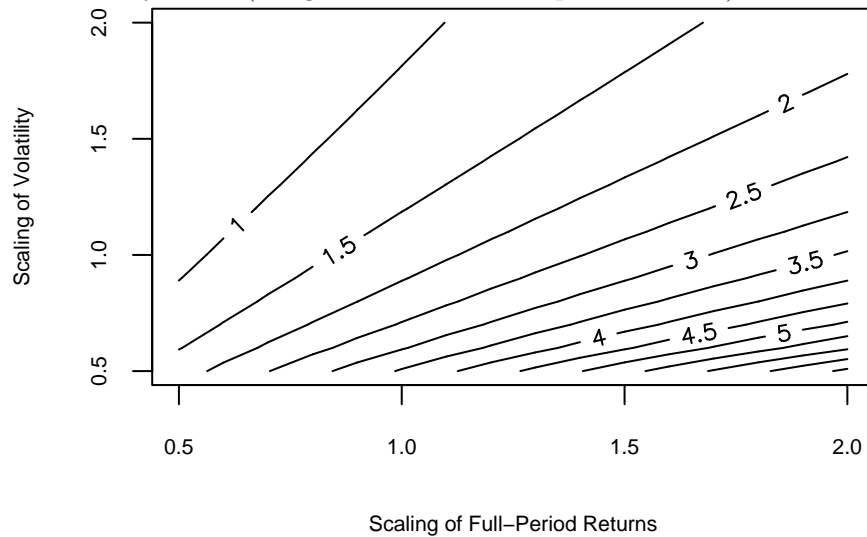


Figure 8: The standard deviation of information ratios as volatility and returns are artificially varied (using data from the first quarter of 2004).



2.3 Measuring Skill via Information Ratios

In order to study the ability to measure skill, a set of 100 “managers” was created. At the beginning of each quarter each manager performs a portfolio optimization. The managers all use the same variance matrix, but each has a unique vector of expected returns. The variance matrix is estimated from the previous two quarters using a statistical factor model. The expected returns in the optimization are based on the actual returns that are realized in the quarter (since this is looking at future data, it is not a strategy that real fund managers have available to them). The expected returns for the stocks are random normals with mean equal to 0.1 times the realized mean daily return for the asset. The standard deviation for the random normals is 0.1 times the standard deviation of the realized daily returns for the asset. The objective of the optimization was to maximize the information ratio—the absolute ratio, not relative to any benchmark.

A common approach to testing for skill is to compute the information ratio of the fund relative to its benchmark. The test is then to see if this information ratio is too large given the null hypothesis that the true value is zero. There are at least two approaches to the test. One is to feed the information ratios for the individual periods—33 quarters in the current case—to a t-test. More common is to calculate the information ratio for the whole period and use the fact that the standard deviation is theoretically known, then use the normal distribution. The statistics and p-values from these two approaches should be similar. Figure 9 shows the p-values from the normal test for the 100 “managers” for the information ratio based on the first random benchmark for the full time period. The skill of the managers shows up by quite a large number having p-values close to zero.

Another view is in Figure 10 which shows the number of hypothetical managers with significant p-values as each quarter is observed—an additional point on the x-axis becomes available as each quarter is completed. There are a couple of aspects to this plot that are worrisome. The number of significant managers is much more variable when only a few quarters have been observed. While the number of managers that are significant at the 5% level grows reasonably steadily as we would expect, the number that are significant at 1% seems to stagnate.

We have seen that the assumption of known standard deviation in the normal test is actually violated. Figure 11 shows the distribution of normal test p-values using information ratios relative to the equally weighted benchmark when there is no skill. Each of the “no skill managers” selects one of the 1000 random portfolios at random each quarter—100,000 such managers were created. If theory were correct, then the distribution in the plot would be uniform (that is, flat). The distribution does not have enough mass in the tails, near 0 and 1. This implies that it is harder (in this case) than it should be to prove fund managers either skilled or unskilled. The deviation from the uniform distribution will be time, benchmark and universe dependent.

Figure 9: P-values of the 100 hypothetical managers based on the information ratio relative to random benchmark 1 over 33 quarters using the normal test.

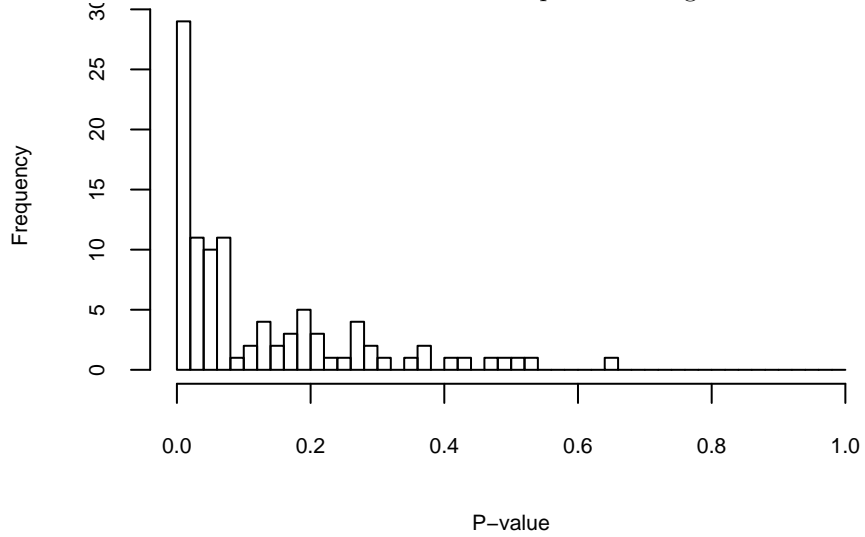


Figure 10: Percent of hypothetical managers with significant p-values from the normal test over time for random benchmark 1.

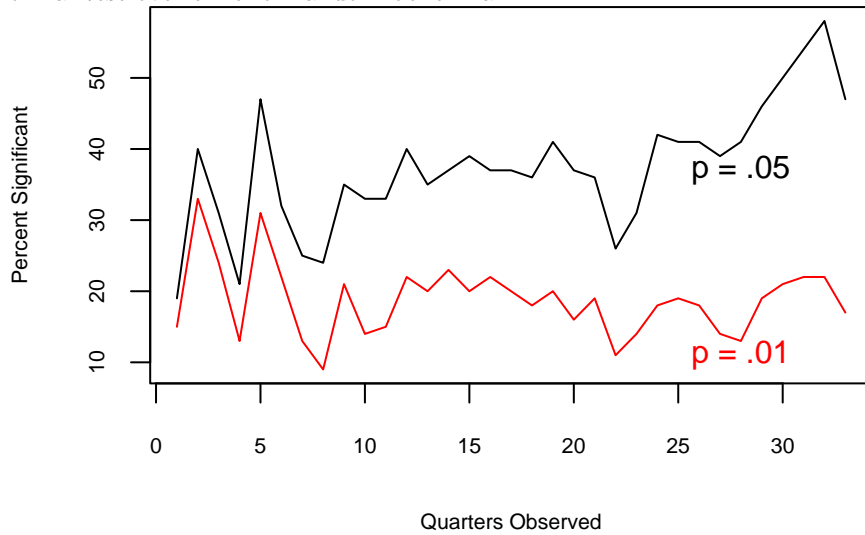
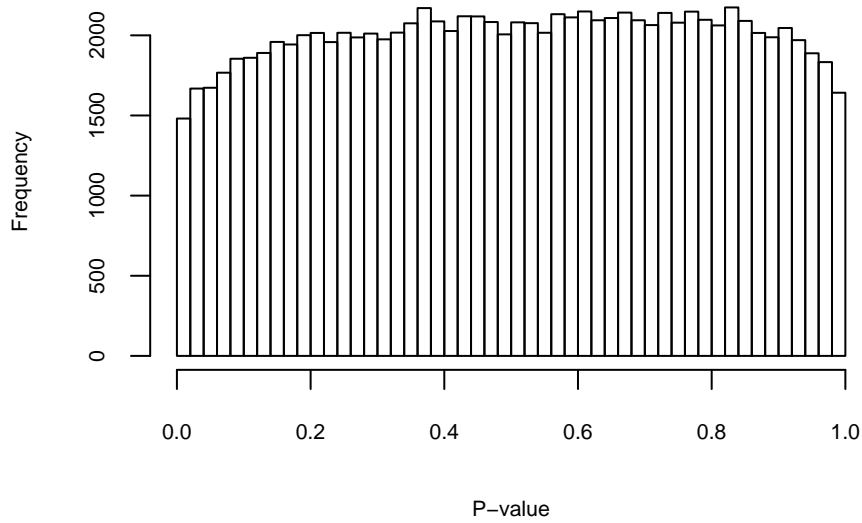


Figure 11: Distribution of p-values from the normal test of information ratios relative to the equally weighted benchmark on portfolios with zero skill.



3 Measuring Manager Skill

We've already seen that assessing the skill of fund managers with information ratios has severe problems.

A second commonly used method is to rank a fund relative to similar funds. This has problems of its own. It supposes that all funds within the category are doing the same thing. For instance, it isn't entirely obvious how differences in volatility should be taken into account, and seemingly small differences in the universe that is used could have a major impact.

Even if all of the funds in a category used precisely the same universe, had the same volatility and so on, we still wouldn't know if the top-ranked managers had skill. It could be that no manager in the category has skill and that the top-ranked managers are merely the luckiest.

Random portfolios provide an opportunity to measure skill more effectively. First, we take a statistical detour.

3.1 Combining p-values

In using random portfolios to measure skill, it will be necessary to combine p-values from different periods of time. A key assumption of the methods of combining p-values that we will explore is that they need to be statistically independent. In our context as long as the tests are for non-overlapping periods of time, this will be true to a practical extent, if not absolutely true.

A formula for combining k statistically independent p-values is

$$P_{combined} = -2 \sum_{i=1}^k \ln(p_i) \sim \chi_{2k}^2 \quad (1)$$

This is called Fisher’s method [Fisher, 1958]. Some of the fame of this method is that it is a good textbook exercise to derive the distribution—see, for example, [Bickel and Doksum, 1977].

An intuitive check on this formula is that an individual p-value of 1 will add 0 to the combined statistic. An individual p-value near 0 will add a large amount to the statistic. If an individual p-value were zero, then the combined statistic would be infinite.

This last possibility brings up the question of what is done when the fund outperforms all of the random portfolios. Naively this seems like a case of a zero p-value, but it is not. The p-value is the probability of seeing a result as extreme or more extreme than what is observed *given the null hypothesis is true*. Therefore the p-value is computed as the number of random portfolios that are as good as or better than the fund plus 1 divided by the number of random portfolios plus 1 (because the observed value counts as well as the random portfolios).

In many cases this adjustment of the p-value that adds 1 to both the numerator and denominator can be considered pedantic. However, when using Fisher’s method of combining p-values, the pedantry is necessary.

Let’s consider data from a particularly erratic fund manager. We have 6 periods, in 3 periods the manager outperforms all 999 random portfolios that we generate and in the other 3 periods the manager underperforms all 999 random portfolios. Thus the p-values for the periods are some permutation of .001, .001, .001, 1, 1, 1. Combining these we get a p-value of about .00004. Thus we are quite convinced that the manager has skill.

However, if we test for negative skill, we get the same individual p-values (in a different order) and hence the same combined p-value. We are in the uncomfortable position of being convinced that the manager both outperforms and underperforms.

One alternative for combining p-values is called Stouffer’s method. In this technique the individual p-values are transformed into the quantiles of a standard normal. The p-value of the average of the quantiles is then found. In R the command to do this is:

```
pnorm(sum(qnorm(x)) / sqrt(length(x)))
```

where x is the vector of individual p-values.

Stouffer’s method easily admits the use of weights for the individual p-values—for example, if not all of the time periods were the same length. A weighted sum of the quantiles is performed, and then standardized by its standard deviation—the square root of the sum of squared weights.

Let's return to our erratic fund manager. Immediately we see a problem because a p-value of 1 will create an infinite value.

When using Stouffer's method, we want to use centered p-values:

$$p_{centered} = \frac{n_x + .5}{N + 1}$$

where N is the number of random portfolios and n_x is the number of portfolios that are as extreme or more extreme than the observed fund.

Using centered p-values for the erratic fund manager, the quantiles of the individual p-values sum to zero, and we get a combined p-value of 0.5 from Stouffer's method. This is much more reasonable than the result from Fisher's method.

Stouffer's method is used to combine p-values in what follows.

3.2 Tests with the Example Data

Figure 9 shows a test of skill using information ratios relative to a benchmark. Here we use the same data to test skill based on the mean-variance utility using random portfolios.

The first step is to decide what specific utility is to be computed. In the case of mean-variance utility we need to specify the risk aversion parameter. We then compute the utility achieved within each quarter by each random portfolio and by each manager. The utility of a manager within a quarter is compared to the utilities of the random portfolios—this provides a p-value for that manager in that quarter. Finally, we combine these p-values to derive a p-value for the whole period for each manager.

Figure 12 plots the p-values based on random portfolio tests using mean-variance utility with risk aversion 2. This has many more very small p-values than Figure 9. Table 2 shows the number of hypothetical managers that achieved various significance levels for different forms of the tests. The tests using random portfolios clearly have more power than those using information ratios. About a third of the random portfolio tests achieve a p-value less than 0.001, while only one manager in one of the information ratio tests achieves this.

Figure 13 shows the number of hypothetical managers with significant p-values as the number of quarters observed increases. This plot shows the number of significant p-values growing rather steadily. The problems that p-values based on information ratios seemed to have are not in evidence in this plot.

4 Investment Mandates

Mandates are the contracts that tell fund managers what they should do with the investor's money. Mandates should be created so that the investor maximizes the usefulness of the entire portfolio. At present this goal is probably not realized very well.

Figure 12: P-values of the 100 hypothetical managers based on random portfolios using mean-variance utility with risk aversion 2 (over 33 quarters).

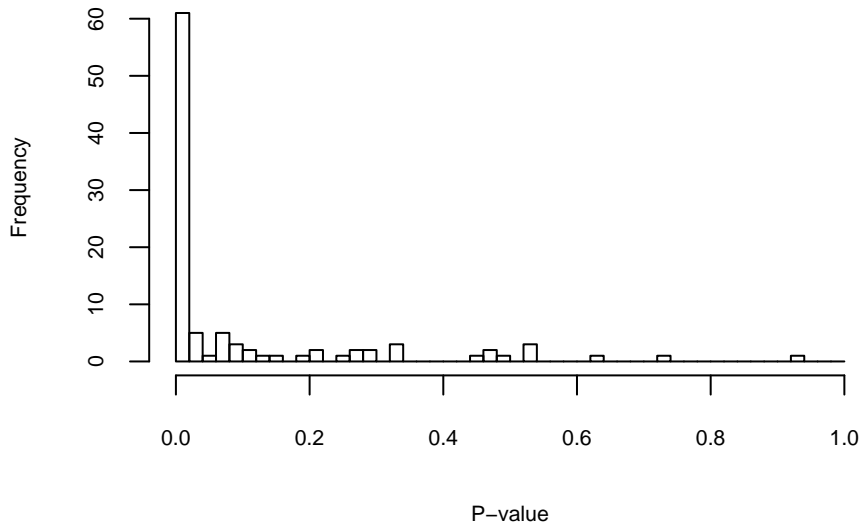
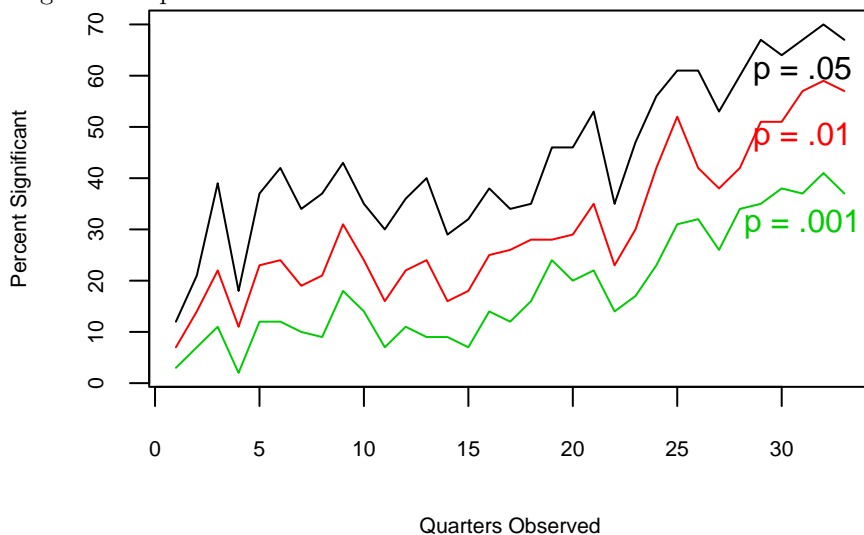


Table 2: Counts (out of 100) of the number of hypothetical managers achieving significance levels in the various forms of tests over 33 quarters.

test	< 0.05	< 0.01	< 0.001
random portfolio, risk aversion=2	67	57	37
random portfolio, risk aversion=1	68	56	35
random portfolio, risk aversion=0.5	67	52	34
random portfolio, risk aversion=0	66	51	33
information ratio, equal wt benchmark	35	12	0
information ratio, random benchmark 1	47	17	1
information ratio, random benchmark 2	43	14	0

Figure 13: Percent of hypothetical managers with significant p-values over time using random portfolios with risk aversion 2.



4.1 Tracking Error Should be Maximized

Currently fund managers are often expected to have a relatively small tracking error to their benchmark. If there were no opportunity to invest passively in the benchmark, then this could be a rational approach. But is this the right approach when passive investment is possible?

If both passive and active funds are held, then the total portfolio is enhanced from lower volatility when the correlation between the passive and active portions decreases (assuming the expected return and volatility of the active fund do not change).

We can see what minimizing correlation means for the tracking error by some minor manipulation of its definition. We will denote the active fund by A and the benchmark by B , other notation should be self-explanatory.

$$TE_B^2(A) = \text{Var}\{A - B\} = \text{Var}\{A\} + \text{Var}\{B\} - 2\text{Cov}\{A, B\} \quad (2)$$

Putting the covariance term alone on the left side and transforming to correlation gives us:

$$\text{Cor}\{A, B\} = \frac{\text{Cov}\{A, B\}}{\sqrt{\text{Var}\{A\}}\sqrt{\text{Var}\{B\}}} = \frac{\text{Var}\{A\} + \text{Var}\{B\} - TE_B^2(A)}{2\sqrt{\text{Var}\{A\}}\sqrt{\text{Var}\{B\}}} \quad (3)$$

Holding the variance of the active fund constant, the correlation between the active fund and the benchmark is minimized when the (squared) tracking error is maximized.

This directly contradicts [Kahn, 2000], cited by [Waring and Siegel, 2003]. Who is right?

4.2 What is Risk?

The argument we've just seen says that tracking errors are ideally large, while [Kahn, 2000] argues that tracking errors should be small. The discrepancy boils down to the definition of risk. The argument in which tracking errors should be large takes the risk to be the mean-variance utility of the entire portfolio—the active part plus the passive part. The argument in which tracking errors should be small takes risk to be the deviation from the benchmark.

Optimal behavior is vastly different depending on which is the more realistic definition of risk.

Calling risk the deviation from the benchmark is the appropriate choice when the benchmark is the liabilities of the fund. If there is no deviation from the benchmark, then the fund carries no risk. For example if the fund needs to deliver x times the value of the S&P 500 in 10 years, then this situation applies with the benchmark equal to the S&P 500.

Alternatively if the benchmark is the S&P 500 but it could reasonably have been some other index of U.S. equities, then exactly reproducing the S&P 500 is not going to be a zero risk solution. This is the more common case.

However, using the absolute utility of the portfolio (where we want to maximize tracking error) is also wrong—it ignores the liabilities altogether, as if we knew nothing about them. As far as I know we don't have the proper mathematics in place to easily evaluate policies when the liabilities are known only with uncertainty. One way of thinking about the problem is that it is a generalization of a dual benchmark optimization. So perhaps an approximate answer can be obtained by performing an optimization with several benchmarks.

My (uneducated) guess is that using the absolute utility is almost always closer to the right answer than using the active utility.

[Muralidhar, 2001] on page 157 speaks of an example where the actively managed portfolio had a lower asset-liability risk (in a certain sense) than the benchmark portfolio. This is obviously a case where deviation from the benchmark should not be considered to be the risk.

Traditionally there has been another reason to prefer small tracking errors: small tracking errors enhance the ability to declare skill when information ratios are used. Consider an extreme case. Two fund managers outperform an index by 3%, their funds each have the same volatility, but one has a tracking error of 1% while the other has a tracking error of 10%. From a global perspective the two fund managers are equivalent—they have the same return and the same volatility. But in terms of proving skill via the information ratio relative to the index, the first fund manager would be judged to have skill while the second could not be.

4.3 Zero-Sum Games

In general active management is a zero-sum game. Active managers try to outperform the average manager. Obviously not all managers can be above average. Thus investors who use active managers need to try to evaluate the quality of the active managers available to them. [Waring and Siegel, 2003] argue that the investor is faced with a portfolio optimization problem where the assets are the fund managers.

The key inputs into this optimization are the expected alphas of the managers. The investor can use random portfolios to help get a sense of these as long as they have returns of the managers and an idea of the constraints that each manager uses. Here the random portfolios would not be used so much to provide a p-value of skill, but rather an estimate of the alpha of a fund manager.

There is a possibility that active management need not be entirely a zero-sum game. If the selection of active managers provides sufficient diversification, then the investor can gain even without an increase in expected return. On page 157 of [Muralidhar, 2001] it says, "... investment teams who are delegated the responsibility of managing the assets should be rewarded for lowering the asset-liability risk even if they do not outperform their benchmarks." This topic is also discussed in [Burns, 2003].

5 Random Portfolio Mandates

Random portfolios can be used as the basis of mandates. The investor specifies the constraints that the fund manager is to obey; the manager is judged, and possibly paid, based on the fund's performance relative to random portfolios that obey the constraints. This process gives fund managers the freedom to shape their portfolios the way that they see fit, and provides investors with an accurate measure of the value to them of a fund manager.

In a traditional mandate the investor and fund manager agree on a benchmark and a tracking error allowance. With a random portfolio mandate, it is the constraints that need to be agreed upon. Of course each party will have views on the constraints.

Fund managers will want the constraints to emphasize their strengths. For example the universe of assets could be limited to some particular set. In general fund managers want constraints to be loose so that they have a lot of freedom, and the random portfolios are allowed to do stupid things.

The investor wants to set the constraints so that the fund manager is likely to add as much value as possible. This tends to favor relatively tight constraints on such things as volatility.

While there is a natural tension between the fund manager and the investor, there is also quite a lot of room for cooperation. It is in the interests of both that the fund manager is given enough freedom to capitalize on good investment ideas.

5.1 An Example Mandate

Here we briefly outline what a random portfolio mandate might look like. The items in our mandate are:

- **The evaluation period is 6 months.**

The frequency of evaluation needs to take the fund manager's strategy into account. Obviously an evaluation over 1 week when the manager is looking at time horizons on the order of 3 to 6 months will be pure noise. A manager that typically holds positions for less than a day could be evaluated very frequently, but the evaluation need not be especially frequent.

- **The universe of assets is the constituents of the S&P 500 at the beginning of the period.**

To keep things simple, the universe is fixed throughout the period regardless of constituent changes in the index itself. An alternative would be to allow new constituents into the universe, in which case the random portfolios would be given the opportunity to trade into the new assets.

- **The number of assets in the portfolio is to be between 50 and 100, inclusive.**

These numbers reflect the desire by the fund manager to hold 100 names or slightly fewer, while the lower bound ensures that the fund never gets too concentrated.

If it is found that the size of the portfolio has a material effect on the distribution of utility, then the random portfolios can be generated with sizes that characterize the actual sizes that the portfolios are likely to be. (In this case the range of allowable sizes would probably be reduced.)

- **The positions are to be long only.**

- **The maximum weight of any asset will be 5%.**

This seems like a straightforward constraint, but isn't—there could be numerous interpretations of what it means. One practical choice is that a position can be no more than 5% at the point when it is created or added to.

- **The volatility of the fund will be no more than 150% of the volatility of the minimum variance portfolio that satisfies the remaining constraints.**

This clearly needs more careful definition. Not just any volatility will do—it has to be agreed. One choice would be to provide a specific variance matrix of the universe of assets. An equivalent approach is to provide the specification of how the variance matrix is to be produced. For example, use the default arguments of the POP function `factor.model.stat` with 4 years of daily log returns.

5.2 Operational Issues

In the example mandate, volatility is constrained statically—only information available at the beginning of the period is used. While this avoids the problem of the fund manager unintentionally breaching the mandate because of changes during the period, it doesn't necessarily state how the investor would like the fund manager to behave. The investor may desire the fund manager to control the volatility of the fund throughout the period using updated information. While slightly more involved, the random portfolios can have trading requirements imposed upon them during the period. However, if the fund manager is being judged based on a utility that includes volatility as a component, then the fund manager should be taking changes in the volatility environment into account in the best interests of the investor.

The evaluation criterion can be at least as useful in shaping the fund manager's behavior as the constraints. The criterion can be anything that can be computed using information that is available at the end of the period—we are not limited to any particular measures such as the return or a mean-variance utility. For example, the criterion might include the skewness of the daily log returns during the evaluation period, and the correlation with some proxy of the rest of the investor's portfolio.

The fund manager may be at a disadvantage (or advantage) relative to the random portfolios if they are allowed unlimited turnover. If a portfolio is already in place, then it is reasonable for the random portfolios to be generated so that there is a maximum amount of trading from the portfolio that exists at the start of the period.

Proposed revisions may arise about the form of the mandate. For example the fund manager may come to think that a particular constraint is not the best approach. With the use of random portfolios the fund manager can demonstrate to the investor the effect of changing the constraint. The mandate can be revised from period to period as more is learned.

5.3 Performance Fees

Performance fees can easily be based on random portfolios from a mandate. As stated earlier, the criterion used to measure success can be specialized to fit the particular situation. As long as the criterion is a close match to the actual utility of the investor, then the interests of the investor and fund manager are aligned when a performance fee is used.

The starting point for a performance fee based on random portfolios is likely to be the average utility of the random portfolios. The fund manager should be paid for utility that is delivered above the base. How much is paid for an increment in utility is, of course, up to the investor. The payment should be for how much utility is delivered, not how difficult it is to deliver.

As in [Waring and Siegel, 2003] the investor is (or should be) doing a portfolio optimization. The investor is selecting weights for a variety of active and passive funds. The investor may assign different utility functions to different

fund managers (for instance the investor could vary what the managers should have a small correlation to), but it is sensible for the investor to pay the same amount to each fund manager for an equivalent increase in utility. More weight should be given to managers who have the ability to deliver a lot of utility.

6 Summary

Random portfolios have been shown to be of use in two respects—measuring skill and forming investment mandates.

The measurement of skill with random portfolios avoids some of the noise that is introduced when performance is measured relative to a benchmark. This means that knowledge of skill can be more precise. An accurate assessment of skill with random portfolios requires a knowledge of both the returns of the fund and the constraints that the fund obeys. Less accurate assessment can be done where the constraints are not known specifically. The statistical statements that result are distribution-free—there are no assumptions on the distribution of returns or measures of utility.

Mandates that are based on random portfolios allow fund managers to play to their strengths because they need not be tied to a benchmark. This also allows more flexibility for the investor to shape the behavior of the fund managers to best advantage.

Other uses of random portfolios include the assessment of the opportunity set available to a fund manager with a given strategy.

References

- [Bickel and Doksum, 1977] Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day.
- [Burns, 2003] Burns, P. (2003). Portfolio sharpening. Working paper, Burns Statistics, <http://www.burns-stat.com/>.
- [Burns Statistics, 2004] Burns Statistics (2004). *POP Portfolio Construction User's Manual*. <http://www.burns-stat.com>.
- [Dawson and Young, 2003] Dawson, R. and Young, R. (2003). Near-uniformly distributed, stochastically generated portfolios. In Satchell, S. and Scowcroft, A., editors, *Advances in Portfolio Construction and Implementation*. Butterworth-Heinemann.
- [Ding and Wermers, 2004] Ding, B. and Wermers, R. (2004). Mutual fund stars: The performance and behavior of U.S. fund managers. Technical report, <http://papers.ssrn.com>.
- [Engstrom, 2004] Engstrom, S. (2004). Does active portfolio management create value? an evaluation of fund managers' decisions. Technical Report 553, Stockholm School of Economics, <http://swopec.hhs.se/hastef/>.

- [Ferson and Khang, 2002] Ferson, W. and Khang, K. (2002). Conditional performance measurement using portfolio weights: Evidence for pension funds. *Journal of Financial Economics*.
- [Fisher, 1958] Fisher, R. A. (1958). *Statistical Methods for Research Workers, 13th Edition*. Hafner Publishing.
- [Grinblatt and Titman, 1993] Grinblatt, M. and Titman, S. (1993). Performance measurement without benchmarks: An examination of mutual fund returns. *Journal of Business*, 66:47–68.
- [Israelsen, 2003] Israelsen, C. L. (2003). Relatively speaking. *Financial Planning Magazine*.
- [Kahn, 2000] Kahn, R. N. (2000). Most pension plans need more enhanced indexing. *Investment Guides, Institutional Investor*.
- [Kosowski et al., 2001] Kosowski, R., Timmermann, A., White, H., and Wermers, R. (2001). Can mutual fund stars really pick stocks? new evidence from a bootstrap analysis. Working paper, <http://papers/ssrn.com>.
- [Muralidhar, 2001] Muralidhar, A. S. (2001). *Innovations in Pension Fund Management*. Stanford University Press.
- [R Development Core Team, 2004] R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, <http://www.r-project.org>. ISBN 3-900051-07-0.
- [Siegel, 2003] Siegel, L. B. (2003). *Benchmarks and Investment Management*. CFA Institute (for hardcopy). Full text available online at <http://www.qwafafew.org/?q=filestore/download/120>.
- [Waring and Siegel, 2003] Waring, M. B. and Siegel, L. B. (2003). The dimensions of active management. *The Journal of Portfolio Management*, 29(3):35–51.