Matching Portfolios

by David Kane and Jeff Enos

April 21, 2008

1 Introduction

"Matching" portfolios is a technique for generating a reasonable benchmark for determining the relative performance of a specific equity portfolio and is based on the work in Ho et al. (2005a). Consider the simplest case of a long-only mutual fund that has returned 10% in the last year. Has the portfolio done well? If the average stock in the universe has gone up 50% then, obviously, the portfolio has done poorly. If, on the other hand, the average stock has gone down 25%, then the portfolio has done remarkably well. In other words, it is impossible to evaluate the performance of a portfolio without considering the hypothetical performance of other possible portfolios. In this case, we are comparing the performance of our portfolio to that of a hypothetical portfolio that is equal-weighted all the stocks in the universe. But, depending on the characteristics of our mutual fund, this may not be a reasonable benchmark.[1]

Matching portfolios provide a benchmark which matches the characteristics — sector exposures, capitalization biases, position sizes — of the target portfolio. The `portfolio` package provides the `matching` method as a means of computing a matching portfolio. In this article we describe the intuition behind matching in general, frame a real-world problem in which computing a portfolio benchmark is difficult, and show how the matching facility of the portfolio package can be used to solve this problem.

2 Data

To focus ideas, let's examine a specific portfolio formed on December 31, 2004. Assay Research[2] is a forensic accounting that provides "in-depth insight into financial statements, accounting practices and policies, and quality-of-earnings of publicly-traded companies." Assay maintains a "Focus List" of companies for which its concerns are "heightened." Although Assay does not provide buy/sell recommendations, most if its customers would expect the stocks on its Focus List to perform poorly going forward.

[1] More information on the performance measurement problem and possible solutions can be found in Burns (2004).

[2] www.assayresearch.com

On December 31, 2004, there were 33 companies on Assay's Focus List. The list of companies, along with data on a total of 4,000 stocks that were trading at that time, are available as part of the R `portfolio` package.

```
> data(assay)
> assay[c(1, 10, 100, 1000, 3000, 767), ]

            id symbol                name country currency price      sector
22101 36960410     GE GENERAL ELECTRIC CO     USA      USD    36 Industrials
29780 47816010    JNJ   JOHNSON & JOHNSON     USA      USD    63     Staples
17157 26054310    DOW        DOW CHEMICAL     USA      USD    50   Materials
57743   663710 NCM.AU NEWCREST MINING LTD     AUS      AUD    17   Materials
52459   646862   8584         JACCS CO LTD     JPN      JPY   616  Financials
6930  G3726010   GRMN          GARMIN LTD     USA      USD    61  Technology
      liquidity on.fl ret.0.3.m ret.0.6.m
22101      2.84 FALSE   -0.0059    -0.021
29780      2.42 FALSE    0.0635     0.078
17157      1.90 FALSE    0.0137    -0.089
57743      0.74 FALSE   -0.0023    -0.115
52459     -0.47 FALSE    0.2138     0.142
6930       1.63  TRUE   -0.2387    -0.297
```

The variables in the data frame follow:

- `id` is an identifier for each security, generally a CUSIP for companies traded on US exchanges and a SEDOL for companies traded elsewhere.
- `symbol` is a human-readable identifier that is generally the ticker that the security trades under in its home market; exchange specific information is sometime appended to it. For example, Newcrest Mining trades under the ticker "NCM" in Australia, indicated by the "AU" suffix.
- `name` is the name of the company. We will generally use the terms "company" and "security" interchangeably even though a "company" is a single legal entity which often has several different types of securities associated with it. In this example, we are only examining the single primary equity security for each company.
- `country` is the ISO code for the country in which the security is traded. Note that this is not necessarily the same as the country in which the company is headquartered or incorporated. For example, Garmin (GRMN) trades in the US but is incorporated in the Cayman Islands.[3]
- `currency` is the ISO code for the currency in which the security trades.
- `price` is the latest closing price for the security as of December 31, 2004. (Not all securities traded on that date.)

[3]http://www.garmin.com

- `sector` is the economic sector in which a majority of the company's business takes place. There are 10 sectors in this data including Communications, Cyclicals, Energy, Financials and Technology.

- `liquidity` is a measure of the typical daily dollar volume of trading in the security. We normalized it to be $N(0, 1)$.

- `assay` is a TRUE/FALSE indicator of whether or not the security was on the Assay Focus List on December 31, 2004. Thirty three companies were on the list at that time.

- `ret.0.3.m` and `ret.0.6.m` are the three and six month, respectively, returns for each security, including dividends.

There are no missing observations. The universe of 4,000 companies consists of large companies and all the AFL stocks, and is restricted to companies that trade on exchanges in developed markets. For example, we include Japan but not South Korea, Austria but not Croatia.

3 An Assay Focus List (AFL) portfolio

Consider a portfolio formed by taking equal-weighted short positions in each of the Assay Focus List stocks and focusing in the returns for the first three months, through March 31, 2005.

```
> assay$assay.wt <- ifelse(assay$on.fl, -1, NA)
> p <- new("portfolioBasic", name = "AFL Portfolio", instant = as.Date("2004-12-31"),
+     data = assay, id.var = "symbol", in.var = "assay.wt", type = "relative",
+     size = "all", ret.var = "ret.0.3.m")
> summary(p)

Portfolio: AFL Portfolio

        count       weight
Long:       0            0
Short:     33           -1

Top/bottom positions by weight:
     id pct
1  ACXM  -3
2  AFFX  -3
3  ANSI  -3
4  ARTC  -3
5    AV  -3
29  SLE  -3
30  UNA  -3
31 USPI  -3
32 UTSI  -3
33  VCI  -3
```

```
> summary(performance(p))

Total return:  7.64 %

Best/Worst performers:
     id weight   ret contrib
18 HOTT -0.030  0.27 -0.0082
21 KOMG -0.030  0.19 -0.0058
2  AFFX -0.030  0.17 -0.0052
26 PDCO -0.030  0.15 -0.0046
12 ELAB -0.030  0.12 -0.0036
24 OPWV -0.030 -0.21  0.0064
17 GRMN -0.030 -0.24  0.0072
3  ANSI -0.030 -0.32  0.0097
5    AV -0.030 -0.32  0.0097
32 UTSI -0.030 -0.51  0.0153
```

This short-only portfolio returns 7.64% because the average Assay stock fell in price by this amount during the first quarter of 2005. Making 7% in three months is rarely a bad thing, but whether or not this counts as good performance depends on how other stocks in the universe performed during this time period. If the universe consistently outperforms the AFL portfolio, then we could achieve better returns by shorting stocks randomly selected from the universe. We might question the utility of subscribing to Assay's focus list examine the opportunity cost of such a subscription.

A simple analysis suggests that the AFL portfolio significantly outperforms the universe. The average stock in the universe of 4,000 was up 1.4%, and the AFL portfolio returned 7.64%. If we shorted 33 randomly selected stocks from the universe, we would expect the return to be -1.4%. If we consider this to be a reasonable benchmark, then the AFL portfolio outperformed the universe by 9%.

But is the Assay portfolio similar to the rest of the universe? To some extent, it is. All the securities in the universe are relatively larger capitalization, liquid equities traded on developed market stock exchanges. But the AFL portfolio is also very different since all of its components are US stocks. Is it fair to use a benchmark with international stocks as a comparison for a US-only portfolio like APL? Probably not.

Another major difference between the AFL portfolio and the universe is that the AFL is concentrated in a limited number of sectors.

```
> exposure(p, exp.var = "sector")

sector
        variable long  short exposure
1 Communications    0 -0.091   -0.091
3    Industrials    0 -0.091   -0.091
2      Cyclicals    0 -0.121   -0.121
4        Staples    0 -0.303   -0.303
5     Technology    0 -0.394   -0.394
```

The analysts at Assay do not place companies from sectors like Financials, Energy and Utilities on to their Focus List because they lack the industry knowledge to evaluate the financial statements for such companies. Any benchmark which includes securities from such sectors is inappropriate for judging the skill of Assay. After all, if Energy stocks did very well in the first quarter of 2005, a benchmark short portfolio which included them would do very poorly. Assay would hardly deserve "credit" for this since it is not claiming that stocks in sectors which it does not cover will do well. It makes no predictions about how energy stocks, on average, will do.[4]

Even if we were to eliminate from the universe both non-US stocks *and* stocks in sectors that Assay does not cover, a variety of incompatibilities between the AFL portfolio and possible benchmarks would remain. For example, the average Assay stock has a liquidity of over 0.5, more than 1/2 a standard deviation greater than the universe as a whole. The Assay portfolio has almost 40% of its holdings in Technology stocks, but the universe is only 7% Technology. What we need is a benchmark portfolio that looks like, that "matches," the Assay portfolio in terms of variables like country, sector, liquidity and so on but which is otherwise randomly selected from non-Assay stocks in the universe.

4 A Matching Portfolio

The solution to the problem of constructing a benchmark for a portfolio like that derived from the Assay Focus List is to create a "matching" portfolio, a portfolio with characteristics as similar as possible to the AFL portfolio without being identical to it. For the AFL benchmark, we would like a portfolio with similar country and sector breakdown as well as a similar distribution of liquidity. If the AFL portfolio does much better (because the Assay stocks do very poorly) than this benchmark, we have evidence that Assay has in fact identified companies with significant problems. It isn't just a matter of the AFL doing well relative to the overall universe because, for example, Energy stocks have risen so much and the AFL isn't short any energy stocks.

4.1 Statistical intuition

One way to think about the assessment of the AFL is to consider an analogy to a randomized scientific experiment. Recall that a randomized experiment or trial begins by selecting a group of subjects to work with. From this population, a group of subjects is randomly selected and to whom is applied the treatment. The rest of the group receives the control. Since the treatment was applied randomly, any differences in the outcome should be the result of the treatment rather than be caused by systematic differences between the treatment and control groups (Rubin (1974)).

Consider a group of 4000 individuals with a headache. We want to determine if the treatment of "taking an aspirin" relieves the headache better than the

[4] Energy was the best performing sector in Q1 2005, with the average stock up over 15%.

control of "taking a placebo." If we only have, say, 33 aspirins to use for the test, we should select 33 people at random from the group of 4000 and give them each an aspirin. The other individuals get the placebo. Afterward, we can see how the treatment group (having taken the aspirin) compares to the control group (who took the placebo). If, for example, the reported headache pain of the treated group is much lower than that of the control group, we might conclude that aspirin works.

Imagine that we have a "treatment" which we believe causes stock prices to fall. We want to test to see if this treatment actually has that effect. The best way to do so is to run a randomized trial. Select, say, 33 stocks at random from the total universe of 4,000 stocks. Apply the treatment to those 33 stocks but not to the other 3,967. If the price of the 33 treated stocks falls more (or rises less) then the prices of the 3,967 control stocks, we might conclude that the treatment works.

The problem arises, for both tests of aspirin and tests of Assay, when we can no longer do random assignment. Imagine that, instead of assigning aspirin/placebo randomly, 33 of the 4,000 people in our group volunteer to take it. The problem is that these 33 might be very different from the others. They might be all men or mostly old or very fat. Unless we somehow "control" for this problem, we will not be able to conclude that the treatment, the aspirin, actually caused the decrease in headache pain. Instead, it could just be that headaches go away more quickly in old, fat men. Instead of comparing our 33 volunteers to everyone else, we need to compare them to a subgroup that "matches" them. If they are mostly male, old and fat, we should select a control group of people who took the placebo that is equally male, old and fat. If aspirin-takers report less pain in this group, then we might reasonably conclude that — at least within this subpart of the population — aspirin works.

The same intuition lies behind the construction of a matching portfolio. We need a portfolio that looks like the AFL portfolio in terms of country, sector and liquidity. If the only difference between the AFL and matching portfolio is that the former consists of stocks that Assay has "heightened" concerns about while the latter consists of similar stocks without such concerns, we may conclude that any differences in their subsequent performance is due to the treatment received.

Now, of course, placement on the Focus List does not *cause* a stock decline in the same way that taking an aspirin causes, by hypothesis, headache pain to decrease. The price of a stock does not decrease as an Assay employee types the focus list, but the price of stocks on the focus list may decrease when Assay customers receive the list and sell or short the stocks on the focus list. Whether information from Assay or conditions internal to a Focus List company cause price decline is not important. What matters is that one can enter a short position before the price declines. Ultimately, the concerns should be first, whether or not Assay can predict negative future performance and second, if one can enter a position before this information is incorporated into the price.

4.2 Making a match

Note: Due to internal changes to the package there may be some inconsistencies in this and subsequent sections of the document.

We want a matching portfolio which is as similar as possible to the AFL portfolio but which does not include the same stocks. The `matching` method in the `portfolio` package provides this functionality, with a little help from the MatchIt package (Ho et al. (2005b)). Calling this method on a `portfolio` object, `p`, returns a portfolio of different stocks that share attributes with the stocks in `p`. These positions most closely resemble those in `p` along the dimensions specified in the `covariates` argument:

```
> p.m <- matching(p, covariates = c("country", "sector", "liquidity"))

> summary(p.m)

Matched portfolio summary: AFL Portfolio

1 matches using greedy matching.

Matched portfolio returns:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.048   0.048   0.048   0.048   0.048   0.048

Original portfolio return: 0.076, with 0 NAs.
Original return relative to matches: 0.029

Original portfolio outperformed 100% of matches.
```

A quick inspection of the new portfolio's positions confirms that none of the positions of `p.m` appear in `p`.

```
> all(!p.m@matches[, 1] %in% p@weights$id)

[1] TRUE
```

Having created a matching portfolio using country, sector, and liquidity as covariates, we would expect `p.m` and `p` to have similar exposures to these variables. First, all of the positions in `p.m` have country `USA`. This makes sense because all AFL stocks, and thus all stocks in `p`, are US stocks. More interestingly, however, the sector exposures of `p.m` are quite similar to the sector exposures of `p`:

```
> exposure(p, exp.var = "sector")

sector
        variable long  short exposure
1 Communications    0 -0.091   -0.091
3    Industrials    0 -0.091   -0.091
```

```
2      Cyclicals    0 -0.121   -0.121
4        Staples    0 -0.303   -0.303
5     Technology    0 -0.394   -0.394
```

```
> exposure(p.m, exp.var = "sector")

$sector
Communications  Conglomerates      Cyclicals         Energy     Financials
        -0.061          0.000         -0.091          0.000          0.000
   Industrials      Materials        Staples     Technology      Utilities
        -0.121          0.000         -0.333         -0.394          0.000
```

The only difference sector-wise between the two portfolios is that `p.m` has one more stock in Staples and one less stock in Communications. Finally, the exposure of `p` to the numeric variable liquidity:

```
> exposure(p, exp.var = "liquidity")

numeric
   variable long short exposure
1 liquidity    0 -0.54    -0.54
```

is quite close to `p.m`'s exposure to liquidity:

```
> exposure(p.m, exp.var = "liquidity")

$liquidity
    1
-0.49
```

Since we matched using more than one covariate, we shouldn't expect the matching portfolio's exposures to the covariates to be exactly the same as those of the original portfolio. However, given a large enough universe upon which the `matching` method can draw, we expect those exposures to be reasonably close.

4.3 The Match as a Benchmark

Now that we've run `matching` on our AFL portfolio and calculated a match, we can examine how the AFL portfolio performed relative to the match. Recall that the AFL portfolio returned 7.64% during Q1 2005, and that members of our 4000 stock universe were up 1.4% on average during this period. The AFL portfolio, then, outperformed a randomly selected portfolio of 33 stocks from our universe by 9%.

The match, however, performed far better than a randomly selected portfolio:

```
> summary(performance(p.m))

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.048   0.048   0.048   0.048   0.048   0.048
```

The match returned 4.78% during Q1 2005, a far better return than the -1.4% of a randomly selected portfolio. If we then use the matching portfolio as the AFL portfolio's benchmark, the AFL portfolio had an excess return of 2.86%. While this excess return is lower than the 9% we would calculate using a randomly selected benchmark, it more accurately reflects the excess return for which Assay should receive "credit".

For example, while the average stock in our universe returned 1.4%, the average US stock returned -3.6%. We could have simply shorted a random collection of 33 US stocks and walked away with 3.6%. Futhermore, stocks in the Technology and Staples sectors on average returned -4.7% and -4.5%, respectively. The matching portfolio, like the AFL portfolio, benefits from having over two-thirds of its positions in these sectors. Finally, stocks with liquidity values close to 0.5, the average liquidity value of AFL stocks, have the same or slightly poorer returns than the average stock in the universe. The AFL portfolio does not perform better or worse than a random portfolio due to its exposure to higher liquidity stocks.

It is clear that the matching portfolio is a better benchmark for the AFL portfolio than a randomly selected portfolio, particularly due to the poor average return of US stocks and stocks in the Technology and Staples sectors relative to the entire universe.

References

Patrick Burns. Performance measurement via random portfolios, 2004. URL `http://www.burns-stat.com/pages/Working/perfmeasrandport.pdf`.

Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart. Matching as nonparametric preprocessing for parametric causal inference, 2005a. URL `http://gking.harvard.edu/files/matchp.pdf`.

Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart. *MatchIt: Nonparametric Preprocessing for Parametric Casual Inference*, 2005b. URL `http://gking.harvard.edu/matchit`. R package version 2.2-5.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.