On Using Statistical Factor Models in Optimizing Long-Only Portfolios

Patrick Burns[*]

6th May 2003

Abstract

Realized tracking errors are examined for a series of optimized portfolios using various estimates for the variance matrix. It is clear that the benchmark should be added mathematically to the variance matrix using the constituent weights—this dramatically outperforms the case where the benchmark is a separate asset in the return matrix or where relative returns are used. The common belief that factor models are to be preferred to sample variance estimates is confirmed, but only on condition that the benchmark is added mathematically to the variance matrix.

1 Introduction

We investigate the estimation of the variance matrix for the purpose of optimizing a long-only portfolio with a benchmark. For simplicity the optimizations merely minimize the tracking error—there are no expected returns given. However, the results should generalize to the case of active portfolios. The conclusions may also be useful for tracking error analyses that do not involve optimization.

An important aspect of the paper is that the realized tracking error is used as the measure of the quality of the optimizations—the usefulness of an optimization is how it actually performs. A similar study in some respects is [3]. However, they do not constrain the number of non-zero weights, and they use a different set of variance estimates.

Sections 2 and 3 are brief discussions of factor models and the handling of outliers. Section 4 describes the data used, and the form of the computations. The results of the computations are given in Section 5. Section 6 summarizes.

---

[*]This report can be found in the working papers section of the Burns Statistics website http://www.burns-stat.com/

2 Factor Model Estimation

It is common in finance to estimate the variance matrix of a set of assets via a factor model. Arbitrage Pricing Theory [6, 5, 2] provides one motivation for using factor models. There are essentially three types of factors that are used for equities. Factors can be fundamental (such as book-to-price or earnings-to-price). Another choice is macroeconomic factors—some discussion of these models is in [1]. The third possibility is statistical factors.

The main advantage of statistical factors is that it is very easy to build the model. The main disadvantage is that there is no clear meaning for the factors. However, a lack of interpretability is not much of a handicap for portfolio optimization.

There are two common approaches to estimating statistical factor models—maximum likelihood (assuming the normal distribution), and principal factors. A rather mathematical treatment of these estimation procedures is in [4]. Maximum likelihood allows a test of whether adding another factor is worthwhile or not. However, you need to worry about the effect of the non-normality of financial returns. Also, maximum likelihood insists that there must be fewer assets than observations.

The principal factor method requires that the modeler choose the number of factors, the initial estimate, and the stopping rule for the iteration. In modeling equities it is common to use the number of factors that account for a specified fraction of the overall variability—this is a computation on the eigenvalues of the sample variance matrix. The number of factors is chosen to be k if the sum of the k largest eigenvalues divided by the sum of all the eigenvalues is approximately equal to the desired fraction. The initial value used in this study was to set the loadings equal to the scaled eigenvectors of the sample variance. Part of the study is to determine a suitable stopping rule for the estimation.

3 Outliers

It is well known that financial returns—especially daily and intraday returns—have longer tails than the normal distribution. This can degrade many estimates that assume the normal distribution. A more serious infraction can occur if a stock split (or similar corporate action) is not accounted for in the price series from which the returns are computed.

There are several approaches to dealing with outliers. A simple method—and the one used in this study—is to Winsorize the data. In words, this pulls outliers in to some maximum distance from the center. The mathematical definition of Winsorization depends on an estimate μ of the central value, an estimate σ of the scale, and a critical value c. Given these, the Winsorization function W is defined as:

$$W(x) = \begin{cases} \mu + c\sigma & x > \mu + c\sigma \\ x & \mu - c\sigma \leq x \leq \mu + c\sigma \\ \mu - c\sigma & x < \mu - c\sigma \end{cases}$$

The estimates used here are the median for μ and the consistently scaled median absolute deviation for σ, known as MAD. The MAD is 1.4826 times the median of the absolute deviations of the observations from μ.

4   Experimental Design

The raw data are daily returns on 200 US stocks. The first return is for 1996-Jan-02, and the last is for 2002-Nov-07. There are no missing values in the data.

The experiments perform a series of optimizations where data up to a specific date are used to create the factor model for the variance matrix, and that optimization is tested on the data for the next 60 trading days. The few runs that require more observations for the variance estimates use a set of 12 dates starting on 1999-Dec-15 and ending 2002-Aug-05.

Portfolios of size 20 and of size 50 are created—part of the optimization is to select which assets are in a particular portfolio. The weights of the assets in the portfolios can be no more than 10%. Each optimization is separate—no constraint is placed on the turnover from one optimization to the next.

The 200 stocks are not a coherent universe—they are an amalgam of large cap and small cap stocks that were at hand. Due to this there is no index that can reasonably be considered to be a benchmark for that universe of stocks. An artificial benchmark was created with the weights that form the minimum variance portfolio using the entire history of returns. Though actual benchmarks do not hold this property, they are likely to be close. Two other benchmarks of random weights were created and used to verify patterns seen with the minimum-variance benchmark. Only results using the minimum-variance benchmark are reported.

Most of the factor models are built with either the previous 250 daily returns, or the previous 100 daily returns. Using only 100 days worth of data is highly unlikely to be a good idea in practice, but it allows us to glimpse the effect of having more assets than observations.

Functions from the POP Portfolio Optimization software of Burns Statistics are used to build the factor models and optimize the portfolios.

5   Results

This section presents the average realized tracking error (in percent) under various choices of building the variance matrix which is used to create a minimum tracking error portfolio. The averaging is over the 24 (or 12) time periods. Tables in this section have columns labelled "nobs" which is the number of observations on which the variance estimates are based. Another column is "size" which is the size of portfolio requested. This is actually the upper bound on the size of the portfolio, but in general the portfolios will be this size.

Table 1: Including, subtracting or adding the benchmark with factor models. The numbers heading the right-most columns are the fraction of variability explained by the factor model.

size	nobs	type	.2	.3	.4	.5	.6	.7	.8	.9
20	250	inclu	7.99	7.73	7.35	6.86	6.19	5.53	5.28	**4.83**
20	250	subtr	5.87	5.13	4.84	4.52	4.37	4.24	4.13	**3.76**
20	250	add	2.20	2.19	2.14	**2.12**	2.14	2.13	2.17	2.19
20	100	inclu	8.67	8.31	8.02	7.66	7.31	6.77	6.19	**5.74**
20	100	subtr	6.53	6.26	5.96	5.87	5.67	5.40	5.14	**5.03**
20	100	add	**2.19**	2.22	2.22	2.25	2.24	2.31	2.31	2.29
50	250	inclu	7.13	6.56	6.07	5.75	5.35	5.01	4.69	**4.25**
50	250	subtr	4.90	4.49	4.19	3.99	3.76	3.55	3.37	**3.04**
50	250	add	1.06	1.06	**1.05**	**1.05**	**1.05**	1.06	1.06	1.06
50	100	inclu	7.56	7.25	6.79	6.45	6.21	5.88	5.53	**5.30**
50	100	subtr	5.73	5.48	5.27	5.23	5.00	4.80	4.53	**4.38**
50	100	add	1.07	**1.06**	1.07	1.07	1.09	1.09	1.09	1.08

5.1 Putting the Benchmark into the Variance

The experiments show that the most important aspect is how the benchmark is incorporated into the variance. Three methods of putting the benchmark into the variance matrix are investigated:

- Include the benchmark as a separate asset in the return matrix from which the variance matrix is formed.

- Compute the variance with returns relative to the benchmark—that is, subtract the benchmark returns from the asset returns. A benchmark is not used in the optimization in this case.

- Add the benchmark mathematically to the estimated variance matrix.

Suppose the variance matrix is V, and w is the vector of the constituent weights (that sum to one) for the benchmark. When the benchmark is added to the variance matrix, Vw is the set of covariances between the benchmark and the assets in V. Hence the variance matrix needs to include all of the constituents of the benchmark—it can contain other assets as well. The variance of the benchmark is computed as $w^T V w$.

Table 1 compares the realized tracking errors of portfolios optimized using the three techniques of incorporating the benchmark. The mathematical adding method dramatically outperforms.

The fraction of variability explained by the factor model has a significant effect, in general. When the benchmark is included in the returns or benchmark returns are subtracted, the results improve as the number of factors increases. The covariance of the benchmark with the other assets are very important, and they will be estimated more accurately as the number of factors

Table 2: Including, subtracting or adding the benchmark with sample variances.

size	nobs	include in returns	subtract	add mathematically
20	250	2.50	2.46	2.27
20	100	3.51	3.63	3.02
50	250	1.82	1.82	1.09
50	100	2.74	2.75	1.71

increases. Clearly the best choice for the fraction of variability explained depends on whether or not the benchmark is added mathematically.

5.2   Factor Model versus Sample Variance

A logical question to ask is if factor models give better answers than sample variances. Given the results in Table 1, another question is the effect of how the benchmark is placed into the variance matrix. Table 2 presents the results when sample variances are used with the benchmark included in the returns, with benchmark returns subtracted from asset returns, and with the benchmark added to the variance matrix.

As with factor models, the results with the sample variances show that adding the benchmark mathematically is much preferred. While all three of the methods will converge to the same answer as the number of observations increases, they can—and obviously do—have different characteristics with limited numbers of observations. Two methods that are equivalent are adding the benchmark mathematically, and using this augmented matrix to create a variance matrix relative to the benchmark.

Comparing the numbers in Tables 1 and 2 shows that factor models perform better than the sample variance when the benchmark is added mathematically. However, factor models perform worse when the other two methods of handling the benchmark are used.

5.3   Fraction of Variability Explained

Given the previous results, the remainder of the experiments always add the benchmark mathematically.

Table 1 suggests that the best choice for the fraction of variabilty explained increases with the number of observations. There is not enough information there to hazard a guess at how fast the increase is or if it asymptotes at less than 100%. In order to get more perspective on this, the shorter set of dates was employed that allows the number of observations used in estimating the variance to be up to 1000. Those results are in Table 3. While there may be a hint that the fraction of variability explained should be increased with the number of observations, there is no clear pattern visible. Undoubtably the small number of time points in the average explains the noisiness of the results.

Table 3: Fraction of variability explained. The numbers heading the right-most columns are the fraction of variability explained by the factor model.

size	nobs	.2	.3	.4	.5	.6	.7	.8	.9
20	1000	2.62	2.60	2.56	2.54	2.53	2.62	2.61	**2.42**
20	500	2.59	2.57	**2.52**	2.57	2.58	2.58	2.60	2.66
20	250	2.59	2.60	2.55	2.58	**2.51**	2.59	2.70	2.69
20	100	2.70	2.70	2.66	**2.59**	2.63	2.75	2.74	2.80
50	1000	1.50	1.50	**1.49**	1.51	1.50	1.50	**1.49**	1.51
50	500	1.50	1.50	**1.49**	1.50	**1.49**	**1.49**	1.50	1.50
50	250	1.50	**1.49**	1.50	1.50	1.50	1.50	1.50	1.50
50	100	1.51	**1.49**	1.50	1.52	1.50	1.51	1.51	1.51

Perhaps what matters is whether or not the number of observations is smaller than the number of assets.

5.4 Principal Factor Iterations

The number of iterations that are optimal to use obviously will depend on the value that is used as the starting solution. Hence the following results only apply when the initial loadings are scaled eigenvectors of the sample variance matrix.

Table 4 shows the effect of changing the number of iterations used to estimate the principal factors. The most visible feature is that the optimizations deteriorate when both the number of iterations and the fraction of variability explained are large—especially when few observations are used to build the model. There seems to be some advantage of one iteration over zero iterations. Allowing more than one iteration may be useful if the fraction of variability explained is chosen strategically. However, it appears there is not much value in trying to optimize the number of iterations precisely.

5.5 The Critical Value of the Winsorization

Table 5 shows the effect of changing the critical value of the Winsorization. There is hardly an effect at all with this data. An effect would surely appear if there were missed splits. The implication is that the long tails of the returns are not exerting a detrimental effect on the factor model estimation.

5.6 Time Weights

It is well known that the true variance matrix of financial data is not constant. A technique that allows for changes while not complicating the modeling is to use time weights. That is, the most recent observations are given more weight than observations farther back in time. The weighting scheme adopted here is to

Table 4: Iterations of the principal factor computation. The numbers heading the right-most columns are the fraction of variability explained by the factor model.

size	nobs	iter	.2	.3	.4	.5	.6	.7	.8	.9
20	250	0	2.22	2.22	2.22	2.16	2.16	2.21	2.27	2.21
20	250	1	2.17	2.11	2.14	2.13	2.13	2.13	2.16	2.14
20	250	10	2.18	2.17	2.10	2.16	2.13	2.20	2.24	2.26
20	250	100	2.23	2.17	2.18	2.15	2.17	2.18	2.23	2.27
20	100	0	2.16	2.15	2.20	2.20	2.22	2.22	2.19	2.27
20	100	1	2.22	2.21	2.21	2.24	2.26	2.33	2.32	2.38
20	100	10	2.14	2.18	2.15	2.25	2.30	2.33	2.57	2.82
20	100	100	2.21	2.21	2.21	2.23	2.30	2.38	2.59	2.76
50	250	0	1.06	1.05	1.06	1.06	1.07	1.06	1.04	1.07
50	250	1	1.06	1.06	1.05	1.05	1.05	1.06	1.06	1.06
50	250	10	1.06	1.05	1.06	1.07	1.07	1.06	1.07	1.10
50	250	100	1.06	1.05	1.06	1.07	1.07	1.06	1.07	1.09
50	100	0	1.05	1.06	1.06	1.05	1.05	1.05	1.05	1.06
50	100	1	1.06	1.06	1.08	1.07	1.07	1.08	1.09	1.09
50	100	10	1.06	1.06	1.06	1.06	1.08	1.12	1.17	1.31
50	100	100	1.06	1.06	1.07	1.08	1.08	1.11	1.13	1.30

Table 5: The critical value for the Winsorization.

critical value	size=20 nobs=250	size=20 nobs=100	size=50 nobs=250	size=50 nobs=100
1	2.12	2.17	1.06	1.07
1.5	2.14	2.17	1.06	1.07
2	2.12	2.20	1.06	1.07
2.5	2.08	2.25	1.05	1.08
3	2.10	2.24	1.06	1.07
3.5	2.10	2.22	1.05	1.06
4	2.14	2.25	1.06	1.07
∞	2.12	2.28	1.06	1.08

Table 6: Time weights with small numbers of observations.

size	nobs	factor weighted	factor unweighted	sample weighted	sample unweighted
20	250	2.11	2.12	2.28	2.25
20	100	2.22	2.23	3.14	3.10
50	250	1.06	1.06	1.09	1.09
50	100	1.07	1.07	1.73	1.65

Table 7: Time weights with an extensive range of the number of observations.

size	nobs	factor weighted	factor unweighted	sample weighted	sample unweighted
20	1000	2.54	2.57	2.48	2.52
20	500	2.57	2.55	2.70	2.64
20	250	2.59	2.57	2.76	2.80
20	100	2.63	2.66	3.68	3.71
50	1000	1.50	1.49	1.51	1.51
50	500	1.49	1.50	1.51	1.50
50	250	1.50	1.50	1.55	1.52
50	100	1.50	1.52	2.12	2.10

give the most ancient observation one-third of the weight that the most recent observation gets, and there is a linear progression between these two extremes.

Table 6 presents the effect of the time weighting versus no time weighting for variance matrices using 100 and 250 observations. There is very little effect. However, this is not unexpected as these are quite short time frames. To explore this further the alternative set of dates was used that allows up to 1000 observations. Those results are given in Table 7.

Table 7 also shows that there is very little effect whether or not time weights are used. It does, however, suggest something else—the sample variance with 1000 observations outperforms the factor model for portfolios of size 20. Closer inspection of the results shows that this is due to the sample variance doing extraordinarily well for the final time period. The factor model outperforms if that time period is dropped. There is a clear pattern of sample variances doing better as the number of observations increases—suggesting that sample variances are particularly vulnerable when the number of assets is larger than or close to the number of observations.

6   Summary

By far the most important effect seen is in how the benchmark is handled. Adding the benchmark mathematically to the variance matrix is vastly superior to the benchmark being an asset in the return matrix from which the variance

matrix is estimated or to using relative returns. Mathematically adding the benchmark will create a matrix that is semi-positive definite. Many optimizers insist that a positive definite matrix is used and will slightly adjust the matrix—this should not have a material effect on results. While the results here are from statistical models, the same advice will surely pertain to fundamental and macroeconomic factor models as well. The key seems to be to have exact covariances for the benchmark given the variance matrix that is being used in the optimization.

This finding is a reasonable explanation for the resistance of many fund managers to portfolio optimization. Someone getting realized tracking errors greater than 3% with a supposedly minimum variance portfolio of 50 stocks is—justifiably—not going to be impressed with the technology. It also suggests that constraints may be doing most of the work in some optimizations that are done.

These experiments support the common view that factor models are to be preferred to sample variances. However, this conventional wisdom appears to be true only when the benchmark is added mathematically to the variance matrix. The advantage of factor models is especially great when the number of assets exceeds the number of observations.

The number of factors used in a model as measured by the fraction of variability explained can have a significant effect. The optimal value is probably driven by a number of variables, but a value close to one-half seems a reasonable default value (if the benchmark is added mathematically to the resulting variance matrix). The effect of the number of factors lessens as the tracking error decreases, so the question of the number of factors is probably more important to active managers than passive managers. The number of iterations in the principal factor algorithm appears to have some impact, but has no clear optimum.

Neither the critical value for Winsorization nor time weights exhibited significance in these experiments. Even so, it is probably best to use these if your computing environment allows them. Winsorization will protect against missed stock splits and other data errors. While not proven useful in this study, time weights are more logical than equally weighting observations.

This research can be extended in several directions. A coherent set of assets with an actual benchmark would provide more realism. It would be especially useful if this were a long history in order to obtain more than the 24 observations available here—our experiments using only 12 time points are particularly noisy. A comparison of maximum likelihood estimation of factor models with principal factors would be interesting. Assets in this study had no missing values. Missing values do appear in reality, and are an important problem in arriving at a good estimate of variance. Long-short portfolios are not addressed here, nor are expected returns (alphas). Yet another direction would be to investigate the amount of bias that can be expected in the optimized tracking error of the portfolio versus the realized tracking error of that optimal portfolio.

References

[1] Patrick Burns. Constructing multinational macroeconomic factor models: Experience from Europe. *Journal of Asset Management*, 1(2):121–131, 2000.

[2] N. Chen, R. Roll, and S. Ross. Economic forces and the stock market. *Journal of Business*, 59:383–403, 1986.

[3] Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraint helps. *Journal of Finance*, 58(4), 2003.

[4] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, 1979.

[5] R. Roll and S. Ross. An empirical investigation of the arbitrage pricing theory. *Journal of Finance*, 35:1073–1103, 1980.

[6] S. Ross. The arbitrage pricing theory of capital asset pricing. *Journal of Economic Theory*, 13:341–360, 1976.