



TI 2009-011/4

Tinbergen Institute Discussion Paper

Robust Optimization of the Equity Momentum Strategy

Arco van Oord¹

Martin Martens²

Herman K. van Dijk³

¹ *Erasmus Research Institute of Management, Erasmus School of Economics, Econometric Institute, Erasmus School of Economics;*

² *Finance Department, Erasmus School of Economics.*

³ *Econometric and Tinbergen Institutes, Erasmus School of Economics.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at
<http://www.tinbergen.nl>.

Robust optimization of the equity momentum strategy

Arco van Oord ^{1*}
Martin Martens ²
Herman K. van Dijk ³

Abstract

Quadratic optimization for asset portfolios often leads to error maximization, with optimizers zooming in on large errors in the predicted inputs, that is, expected returns and risks. The consequence in most cases is a poor real-time performance. In this paper we show how to improve real-time performance of the popular equity momentum strategy with robust optimization in an empirical application involving 1500-2500 US stocks over the period 1963-2006. We also show that popular procedures like Bayes-Stein estimated expected returns, shrinking the covariance matrix and adding weight constraints fail in such a practical case.

¹ Erasmus Research Institute of Management, Erasmus School of Economics, Econometric Institute, Erasmus School of Economics; people.few.eur.nl/vanoord

² Finance Department, Erasmus School of Economics

³ Econometric and Tinbergen Institutes, Erasmus School of Economics; people.few.eur.nl/hkvandijk

* Corresponding author: vanoord@ese.eur.nl (e-mail), +31 104088908 (tel.)

1. Introduction

Markowitz' portfolio theory implies that investors should carefully choose stock portfolio weights to optimally balance return and risk, resulting in a quadratic programming problem. In practice portfolio optimization is often known as error maximization. Reason is that standard portfolio optimization magnifies the errors in the return and risk input. Several solutions are proposed in the literature to address this problem. First, predicted inputs (expected return and risk) can be modified to reduce the impact of potential errors. Examples are the Bayes-Stein estimator to improve upon the sample mean (Stein, 1955) and using a weighted average of two covariance matrices (Ledoit and Wolf, 2003; 2004). Second, Jaganathan and Ma (2003) show the merits of imposing weight constraints. Third, Michaud (1998) introduced portfolio re-sampling. Fourth, robust portfolio optimization (Goldfarb and Iyengar, 2003) explicitly deals with the issue of estimation errors in the inputs.

A major drawback in the literature is that for expected returns often long-term historical averages are used, which have little explanatory power for future returns. Hence any solution that shrinks these poor expected return estimates would almost always look good. In this study we do consider a useful input for expected returns, known as momentum. The equity price momentum strategy introduced by Jegadeesh and Titman (1993) buys the top 10% (winners) and sells the bottom 10% (losers) of the past 6 months, resulting in an equally weighted zero-investment portfolio. For US stocks the returns to this zero-investment strategy averaged 6.5% per annum from 1926 to 2005. Its simplicity and its success is one of the most intriguing and as of yet unexplained anomalies that violate the efficient market hypothesis. We note that this result is achieved with a very simple portfolio construction rule that ignores risk. Martens and Van Oord (2008) show, for instance, that through time the momentum strategy has significant time-varying risk exposures to the well-known Fama and French (1993) risk-factors, i.e. market, size and value-growth. Still, as DeMiguel et al. (2007) show, defeating the equally weighted strategy is quite a difficult task.

Another challenging feature of the momentum strategy is the sheer size of the universe. Most existing studies focus on relatively small universes, whereas here we have 1500 to 2500 stocks to choose from each month. This provides more opportunities for error maximization and hence poses an interesting challenge. Furthermore the momentum strategy is a zero-investment strategy, with a long and short portfolio of equal size. We will show that in this setting robust optimization effectively chooses a different risk-return trade-off whereas the Bayes-Stein estimator also does not affect the optimal holdings and thus does not manage to reduce error maximization.

In this study we try to improve upon the simple long-short momentum portfolio using quadratic optimization. To do so we need to overcome error maximization in the setting of a large universe and a zero-investment portfolio. We start with the raw momentum signals, measured by the past 6-month return. Next we address the mapping of the raw signals into expected returns, the estimation of the covariance matrix, and finding the optimal portfolio weights. We do so both by considering Bayesian shrinkage estimates of the parameters needed for the expected returns and covariance matrix, and by applying useful weight constraints. Doing this with a known useful return input is a good test of the adequacy of different techniques aimed at avoiding error maximization. Both the top-bottom 10% strategy and standard quadratic programming without input modification are benchmarks with a good performance and hence tough to beat, in contrast to benchmarks used by existing studies. In the process we also show the added value of portfolio optimization over heuristic rules

provided one can properly deal with uncertainty in the input. This is done for real-world data with practical relevance as opposed to a controlled simulation environment. Finally we explain all the difficulties we had to overcome and the solutions we came up with to create a convex quadratic optimization problem for this practical case.

The outline of this paper is as follows: Section 2 describes our methodology and in particular how to apply popular robustification techniques to large long-short portfolios, Section 3 describes the data and the performance of the standard momentum strategy, Section 4 discusses the results of all suggested portfolio robustifications and Section 5 concludes and provides suggestions for further research towards robust optimization techniques.

2. Methodology

2.1 Objective function and weight constraints

Our benchmark is the standard momentum strategy. In this strategy at the end of each month $t-1$ all stocks are ranked based on their return over the 6-month period from $t-7$ to $t-2$. Subsequently the strategy invests 100% in the top decile (the past winners) and finances this by shorting the bottom decile (the past losers). Next, the month $t-1$ is skipped because in the literature (see e.g. Jegadeesh and Titman, 1995) it is found that losers outperform winners in this month. The long and short portfolios are equally weighted.

Here we also want to construct a portfolio that is 100% long and 100% short. In our case, however, we want to employ quadratic optimization to optimize the trade-off between risk and return. In particular we have the following objective function:

$$\begin{aligned} & \text{Max } h_t' f_t - \lambda \cdot h_t' V_t h_t \\ & \text{s.t. } h_t' \mathbf{1} = 0 \quad (\text{Zero investment}) \\ & \quad \sum_{h_{t,n} > 0} h_{t,n} = 1 \quad (\text{100\% long, 100\% short}) \end{aligned} \tag{1}$$

Each month t we can choose from N stocks. Here h_t are the N holdings (stock weights) we need to choose, f_t is the $(N \times 1)$ vector of expected returns for each stock, λ is the risk-averseness parameter which measures the trade-off between portfolio return and portfolio variance, and V_t is the $(N \times N)$ expected co-variance matrix. The two weight restrictions are imposed to make sure we get a zero investment portfolio by going 100% long and 100% short like the standard momentum strategy. Without this constraint mean-variance might simply apply leverage to the standard momentum strategy to achieve higher returns, whilst not increasing the Sharpe ratio (the ratio of return and risk).

Optimization problem is not convex

The problem with equation (1) is that it is a non-convex problem due to the final restriction. Leaving out this restriction and afterwards scaling the weights of the resulting optimal portfolio to satisfy the 100% long and 100% short restriction makes λ redundant: there is just one ex-ante maximum Sharpe ratio portfolio. We do, however, want to steer the optimizer away from just a single portfolio, as the ex-ante maximum Sharpe-ratio portfolio is also the portfolio that is most likely to maximize the errors in both the expected returns and the covariance matrix.

Introducing two new vectors: longs and shorts

A solution to the problem that the objective function in (1) is not convex is to introduce two new ($N \times 1$) vectors, one for the long positions (l), and one for the short positions (s), such that $l + s = h$. For each of the two vectors it can be imposed that the sum of its elements adds up to one, hence imposing 100% long and 100% short. There is, however, a catch. The solution is still the same ex-ante maximum Sharpe ratio portfolio, because the 100% long and 100% short restriction can be circumvented easily: For the final solution simply put the same additional weights in l and s , which cancel out and hence have no effect on the Sharpe ratio, but do make it satisfy the constraints.

Another weight constraint

To counter the aforementioned problem of opposing weights in the same stock in the ‘long’ and the ‘short’ weight vectors we impose the constraint that a stock can have only long weights if its expected return is in the top 50%, and only short weights if its expected return is in the bottom 50% of the expected return for the coming month. It is in this setting that we really target a 100% long and 100% portfolio, and have a convex optimization problem. Furthermore we also avoid extreme long and short weights in stocks that are similarly attractive in terms of expected return, simply to off-set risks. Hence this also helps to reduce error maximization.

2.2 Expected returns

As mentioned before a unique feature of this paper is that we use well-known useful information regarding expected or predicted returns. Most studies that focus on avoiding error maximization use long-term averages that have little predictive power. In our case the starting point for expected returns is the sum of the past 6 months of returns, respecting a skip month,

$$f_{t,n} = \sum_{\tau=-7}^{\tau=-2} r_{t-\tau,n} \quad (2)$$

where $r_{t,n}$ is the return on stock n in month t .

Using equation (2) will lead to a large dispersion in expected stock returns, resulting in the risk of error maximization as the optimizer may put a lot of weight in the stocks with the highest and lowest expected returns. A popular approach to get more robust predictions of the expected returns is Stein's (1955) estimation procedure for means as proposed by Jorion (1986) for portfolio optimization. Jorion (1986) states that the expected returns are given by

$$f_{t,n}^{BS} = (1-w)f_{t,n} + wf_t^{MV} \quad (3)$$
$$f_t^{MV} = \left(\frac{t' \Sigma^{-1}}{t' \Sigma^{-1} t} \right) f_t$$

$$w = \frac{N+2}{N+2 + (f - f_t^{MV} t)' T \Sigma^{-1} (f - f_t^{MV} t)}$$

where t is a vector of ones and Σ^{-1} has to be replaced by the inverse of the sample covariance matrix. This raises the first problem: the sample covariance matrix does not have an inverse if the number of assets N is too large in comparison to the return history T . The Bayes-Stein estimator of expected returns f^{BS} is a combination of the sample mean (in this case computed

following (2)) and the expected return of the mean-variance combination f^{MV} . For a larger weight w the expected returns for the different stocks become closer to each other; less dispersion in the expected returns results in portfolios closer to the minimum variance portfolio. In this momentum implementation T is smaller than in general holds for historical averages. Therefore the momentum signals f in (2) are shrunk more towards the minimum variance expected return f^{MV} than historical averages. We argue that this is an unfavorable result, since we use the momentum signals just because it has more information on future expected returns than long-term historical averages. Most importantly, because we have a zero-investment problem, the expected return of the portfolio is not affected by the f^{MV} term. Hence this leaves dividing all expected returns by the same number, which is equivalent to changing the risk-averseness parameter in equation (1). Hence for our problem the Bayes-Stein estimator does not help at all to reduce error maximization.

We have, however, implemented four pragmatic robustifications of the expected returns in equation (2). In all these cases we impose that large groups of stocks have the same expected returns, reducing the danger of error maximization. First, we use the average return of the momentum deciles over the entire sample period although we know we are using future information in that case. Second, we use a feasible variant by using the average return of the momentum deciles with an expanding window from the start of the sample period to the current month. Third, we weight the feasible variant with the raw expected returns in (2) as we expect that the group averages do not profit from all information available in the raw expected returns. Finally we use the averages over 20 buckets (vigintiles) instead of over 10 buckets (deciles) for both the feasible and infeasible variant.

The expected returns are all standardized before they are used in the optimizer for a good comparison for equal risk-aversions over methods with different expected returns. The location (cross-sectional average) of the expected returns f_t does not affect the optimal holdings for long-short portfolios, while changing the scaling (cross-sectional standard deviation) of the expected returns is equivalent to changing the risk-aversion in (1).

2.3 Covariance matrix

Jaganathan and Ma (2003) and Fan et al. (2008), among others, show that structural models lead to more robust covariance matrices. Furthermore, we have 1500 to 2500 US stocks in our dataset, hence the sample covariance matrix is not even positive definite if we look back using the popular choice of 60 months. Finally, Martens and Van Oord (2008) show that through time the momentum strategy has significant time-varying risk exposures to the well-known Fama and French (1993) risk factors, i.e. market, size and value-growth. We therefore use the following 3-factor model for the stock returns:

$$r_{t,n} = \alpha_n + \beta_n r_{Mt} + s_n SMB_t + \delta_n HML_t + \varepsilon_{t,n} \quad (4)$$

where $r_{M,t}$ is the market return in month t , SMB_t is a long-short portfolio that is long the smallest market capitalization stocks and short the biggest market capitalization stocks, and HML_t is a long-short portfolio that is long value stocks (high book-to-market ratio) and short growth stocks (low book-to-market ratio)⁴. At the end of each month we estimate equation (4) for each available stock based on data for the past 60 months, with a minimum requirement of 24 available months. This results in estimates for β_n , s_n and δ_n . Let X_t be the $(N \times 3)$ matrix containing all the estimated coefficients using data up to month $t-1$, and let F_t be the sample

⁴ We would like to thank Kenneth French for making these data available on his website.

covariance matrix of the three factor returns using the past 60 months. Then we use for the covariance matrix,

$$V_t^F = X_t F_t X_t' + \Delta_t \quad (5)$$

where Δ_t is a $(N \times N)$ diagonal matrix with the residual variances, i.e. on the diagonal of V_t we get the sample variances based on the past 60 months.

We also consider two ways to make the covariance matrix more robust. First we can use shrinkage on the factor exposures. Therefore we assume a hierarchical model for the risk model, in particular we assume a flat prior in the random coefficient model for the parameters in (4). This risk model becomes

$$\begin{aligned} r_{t,n} &= \alpha_n + R_t^F \beta_n + \varepsilon_{t,n} \quad \varepsilon_{t,n} \sim N(0, \sigma_\varepsilon^2) \\ \alpha_n &\sim N(\alpha, \sigma_\alpha^2) \quad \beta_n \sim N(\beta, \Sigma_\beta) \end{aligned} \quad (6)$$

where β_n contains the three factor exposures of stock n . We assume flat priors for all parameters σ_ε , α , σ_α , β and Σ_β . The final shrinkage estimates for the factor exposures are found by either maximizing a highly nonlinear likelihood or Gibbs sampling. Our large portfolio optimizations are already very time-consuming and therefore we choose to use an intuitive shrinkage

$$\beta_n^S = \left[(X'X) / \sigma_\varepsilon + \Sigma_\beta^{-1} \right]^{-1} \left((X'X) \beta_n / \sigma_\varepsilon + \Sigma_\beta^{-1} \bar{\beta} \right) \quad (7)$$

where σ_ε is the standard error of the OLS regression of (4), β_n is estimated in (4) using OLS, Σ_β the cross-sectional covariance matrix of all these estimated OLS betas and $\bar{\beta}$ is the cross-sectional average of the OLS betas. This results in a large shrinkage towards the cross-sectional beta for stocks with a large standard error of regression, the stocks that have most likely erroneous betas. If a stock's betas are close to the cross-sectional average, but the uncertainty is large, its betas do not change a lot of course. Thus, betas with a large estimation error are shrunk more towards the cross-sectional average than stocks with a low standard error of regression.

Second we can combine the covariance matrix in equation (5) with the sample covariance matrix, see Ledoit and Wolff (2003). The covariance matrix is then of the form

$$V_t^{LW} = w V_t^F + (1 - w) V_t^{Hist} \quad (8)$$

where V_t^{Hist} is the historical covariance matrix over the past 60 months. It is very time-consuming to determine the weights according to Ledoit & Wolf (2003), since we have to deal with a very large covariance matrix; every month we would have to sum and multiply the return histories of more than 1500 stocks. Rather than computing the optimal shrinkage factors we pick 0.99, 0.95 and 0.90 as weights on the factor covariance matrix for the sake of computation time. Ledoit & Wolf (2003) find an average weight of 0.80 on the shrinkage target if they use 120 months of history and have a maximum of 1200 stocks in their universe. We use a higher weight on the shrinkage target as we have more stocks and use less history which both increase the weight on the shrinkage target.

2.4 Weight constraints

Besides the weight constraints already mentioned in Section 2.1 we also test the merits of imposing a maximum weight constraint for each individual stock as also is done by e.g. Chan et al. (1999) and Jaganathan and Ma (2003). There the target is a long-only portfolio and the non-negative constraint is very important, but also here it will avoid extreme weights in individual stocks.

3. Data and equity momentum

Total monthly returns of all NYSE and AMEX stocks in the CRSP database are used, covering the period 1926 to 2005. Stocks that have a stock price lower than \$5 at any moment during the ranking period are excluded as also done by Amihud (2002) and Pastor and Stambaugh (2003). We use a six-month ranking period, wait for one month to avoid price reversals⁵, and then go long in the 10% winners and short in the 10% losers for one month. We rule out those stocks that do not exist in the whole six-month ranking period as well as stocks that do not have at least 24 months of history for the estimation of the risk models. The number of stocks available per month shows a sharp increase in August 1962; we therefore choose to start in this month as our focus is on large long-short portfolios.

Table I shows the sample characteristics of the standard momentum strategy. The first two columns show the characteristics of the standard momentum strategy, while the last two columns show these characteristics if the momentum strategy uses the top and bottom 5% (vigintiles) rather than the top and bottom deciles of the ranking. The first and third column show these results if only stocks are included that only need to exist during the 6 months of the formation period, whereas the second and fourth column only use stocks that also exist during the previous 24 months to ensure a proper estimation of the risk model.

Table I: Annualized characteristics of standard momentum strategy

	I	II	III	IV
Top/Bottom	deciles	deciles	vigintiles	vigintiles
Required history	6 months	24 months	6 months	24 months
Average return	0.0769	0.0719	0.1065	0.1087
Volatility	0.1582	0.1518	0.1882	0.1793
Information Ratio	0.4859	0.4734	0.5658	0.6064

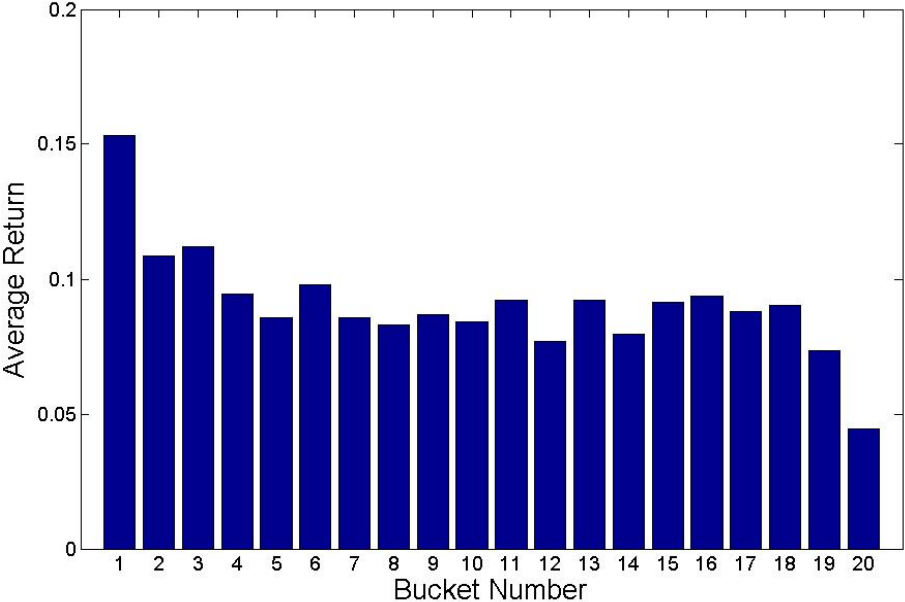
Note: This table shows the annualized return, realized volatility and information ratio of the momentum strategy starting in February 1963 (August 1964) until December 2005 in columns I and III (II and IV) if the stock universe consists of stocks that exist for at least 6 (24) months. The strategies reported in columns I and II (III and IV) buy the top decile (vigintile) of the 6 month ranking and short-sell the bottom 10% (5%) of this ranking after a skip month.

Key in understanding our momentum results is the difference between the forecasted and realized returns of the deciles or vigintiles of the momentum strategy. Figures I and II show the average realized and forecasted returns (using equation (2)) of the vigintiles, respectively, if we require that the stocks exist for at least 24 months. The forecasted vigintile returns in Figure II show a nice decreasing pattern from the top to the bottom vigintile. The realized vigintile returns in Figure I however, only show this pattern for the top and two bottom vigintiles; the remaining vigintiles are not strictly decreasing. This introduces problems in the portfolio optimization process when using equation (2) for the expected returns. The optimizer finds stocks in vigintile 2 more attractive than stocks in vigintile 3, while ex-post vigintile 3 outperforms vigintile 2. Besides this difference between the forecasted and realized

⁵ See e.g. Jegadeesh and Titman (1993, 1995), Lehmann (1990) and Lo and MacKinlay (1990).

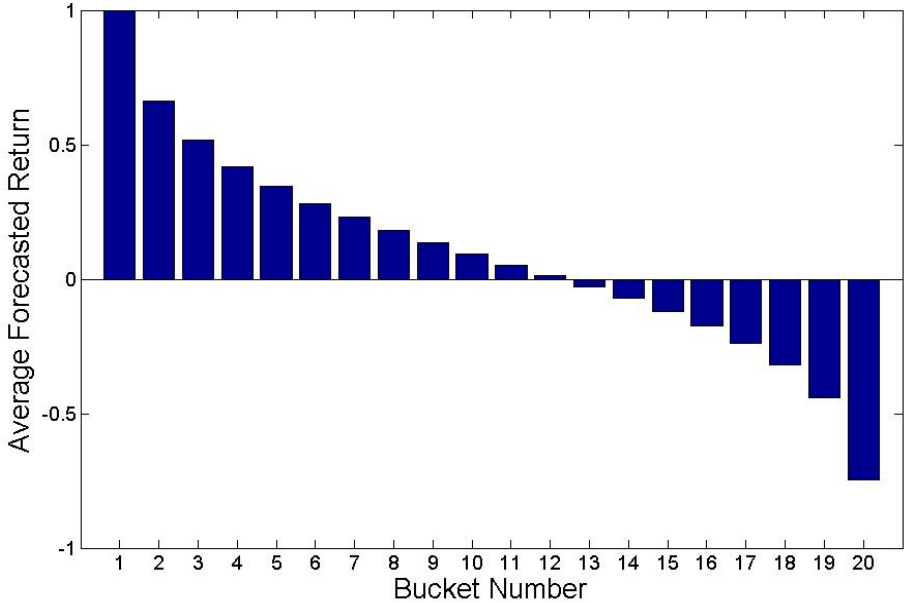
bucket returns, there is also a large difference in the scale of the returns. The forecasted bucket returns, the realized past six month average over all stocks in a particular bucket, range from plus 100% for the top vigintile to minus 80% for the bottom vigintile.

Figure I: Realized vigintile returns



Note: This figure shows the realized vigintile returns per annum from August 1964 until December 2005 if we sort all stocks that exist at least 24 months on their performance in the last six months, subsequently skip one month and then invest in the $[(Bucket\ number - 1) \times 5\%]$ till $[(Bucket\ number \times 5\%)]$ of the ranked stocks.

Figure II: Forecasted vigintile returns



Note: This figure shows the average forecasted return per annum according to equation (2) for the period August 1964 until December 2005 of stocks that existed at least 24 months. The forecasts for vigintile *Bucket number i* are averaged over all stocks that were in $[(Bucket\ number\ i - 1) \times 5\%]$ till $[(Bucket\ number\ i \times 5\%)]$ of the ranking on the past six months returns after the skip month.

4. Results

We have solved the optimization problem in equation (1) for several values for the risk-averseness parameter, λ : 100, 500, 1000, 5000 and 10000. These risk-aversions are uncommonly high compared to risk-aversions found in other research, but standardizing the expected returns leads to higher absolute expected returns that we have to correct for with these higher risk-aversions as discussed in Subsection 2.2. Usually expected returns range from -20% to 20%, whereas 95% of the standardized expected returns ranges from -196% to 196%. A low risk-aversion λ indicates more emphasis on the expected returns estimation of f_t , while a high risk-aversion moves the emphasis to the estimation of the covariance matrix V_t . Consequently we expect that the robustification of the expected returns (covariance matrix) will increase the ex-post Sharpe-ratio for lower (higher) risk-aversions. In general we expect that the robustifications do not always increase the maximum realized Sharpe-ratio, but will improve the Sharpe ratios given a particular low or high level of risk-averseness.

4.1 Impact of using different return inputs

We start with showing the results for different ways to compute the expected return based on momentum returns. For this we use the Fama and French 3-factor covariance matrix, as we know factor models are robust (see e.g. Fan et al., 2008) and the market return, size return and value-growth premium are important risk dimensions for the momentum strategy, see Martens and Van Oord (2008).

Table II shows the results. First we use the return from month $t-7$ to $t-2$ for each stock, see equation (2). In this case we use a different expected return for each stock, at the risk that the most extreme past returns are in no relation to the subsequent realized next month's return. Of all methods this is the approach that is most prone to error maximization. The best Sharpe ratio of 0.531 is quite good. When we choose, however, a low level for λ meaning we focus more on the expected returns than on the expected risk, we see that the Sharpe ratio is quite low at 0.106. Effectively, the optimizer just buys a few top-listed stocks and short-sells the bottom ranked stocks. For lower risk-aversions there is a strong emphasis of the optimization on the expected returns; the optimizer chooses to invest in only the best stocks according to the momentum ranking and thus suffers from a lack of diversification.

Second we use two solutions where the expected return is made more robust by requiring it to be the same for all stocks in the same decile or vigintile, where for each decile and vigintile we use the average ex-post returns of each bucket (see Figure 1 for the vigintiles). For a forecaster standing in 1965 this information would be unavailable, but so of course is the notion that momentum is a good strategy in the first place. In columns II and III we can see that the Sharpe improves when going from 10 to 20 buckets, hence there is information in providing more detailed information. Also we see that for 20 buckets the best Sharpe ratio is 0.646, substantially higher than the 0.531 for raw returns. Furthermore the outcome is much less sensitive to the risk averseness parameter.

Finally we use two solutions that are the same as the one just discussed, but for using only information up to month t where you stand each time when running the optimizer to determine the expected return of each bucket. Again we see that using 20 buckets is better than using just 10 buckets and that the foresight information underlying the best Sharpe in column III at 0.646 is run close by the one in column V at 0.618 and both are a substantial improvement over the raw returns from equation (2) as input for expected returns.

Columns II-IV in table II do not show a lot of dispersion over the different risk-aversions. This is explained by the fact that the optimizer only chooses stocks from the top and bottom decile (vigintile). These chosen stocks all have the same expected bucket return and the optimizer thus minimizes the variance over these selected stocks. This yields almost equal portfolios over the different risk-aversions. This also explains why the standard momentum strategy for vigintiles as presented in Table I column IV and the mean-variance implementation lead to similar results. Both methods simply go long in the top vigintile and short-sell stocks from the bottom vigintile. It appears that diversification according to the Fama & French 3 factor risk model does not add much value over simply equally weighting the stocks in the top and bottom vigintile. We do see, however, that when choosing the highest risk-averseness parameter in the final row, that the more accurate expected return with 20 buckets (i.e. based on the full sample instead of only the in-sample) is important when the focus is on risk, because in that case the stocks in buckets 3, 4 etceteras matter, and it is exactly for these that the full-sample provides a more accurate picture.

Table II: Optimization of the momentum strategy with different expected returns

	I	II	III	IV	V
Returns	Raw returns	10 buckets (inf)	20 buckets (inf)	10 buckets (feas)	20 buckets (feas)
Covariances	F&F Cov	F&F Cov	F&F Cov	F&F Cov	F&F Cov
Constraints	None	None	None	None	None
$\lambda_1 = 100$	0.106	0.469	0.617	0.469	0.617
$\lambda_2 = 500$	0.314	0.469	0.617	0.469	0.617
$\lambda_3 = 1000$	0.418	0.469	0.620	0.469	0.618
$\lambda_4 = 5000$	0.531	0.498	0.646	0.462	0.586
$\lambda_5 = 10000$	0.487	0.500	0.631	0.454	0.566

Note: We solve the quadratic optimization problem in equation (1) for different choices for the risk-averseness parameter λ , different choices for the expected returns f_t , and the Fama and French 3-factor covariance matrix in equation (5). The choices for the expected returns are the return from month t-7 to t-2 for each stock (see equation (2)) in column I, the average ex-post return of the momentum deciles or vigintiles from 1965-2006 in columns II and III, and the average ex-post return of the momentum deciles or vigintiles from 1965 to the month t-1 to predict the next month t in columns IV and V. The table shows the Sharpe ratios based on the realized (ex-post) monthly returns and standard deviation of these returns.

In Table III we show results when combining the raw return inputs from equation (2) with the expected returns from the bucket approach. Here we put much more weight on the bucket returns because these are much less extreme for the top and bottom buckets than the outcomes from equation (2). The results show that even with a small weight for the volatile returns from equation (2) we cannot improve upon the results in Table II for just basing expected returns on the buckets. Only in column II we see an outcome of 0.534 that is slightly higher than using only expected returns from equation (1) that results in a Sharpe of 0.531.

Table III: Combined (raw & bucket feasible) expected returns

	I	II	III	IV
Returns	10 buckets (0.05)	10 buckets (0.10)	20 buckets (0.05)	20 buckets (0.10)
Covariances	F&F Cov	F&F Cov	F&F Cov	F&F Cov
Constraints	None	None	None	None
$\lambda_1 = 100$	0.179	0.135	0.185	0.139
$\lambda_2 = 500$	0.415	0.376	0.425	0.384
$\lambda_3 = 1000$	0.492	0.460	0.517	0.476
$\lambda_4 = 5000$	0.522	0.534	0.574	0.565
$\lambda_5 = 10000$	0.484	0.491	0.538	0.525

Note: We solve the quadratic optimization problem in equation (1) for different choices for the risk-averseness parameter λ , different choices for the expected returns f_t , and the Fama and French 3-factor covariance matrix in equation (5). The choices for the expected returns are a weighted average of the return from month $t-7$ to $t-2$ for each stock (see equation (2)) and the average ex-post return of the momentum deciles or vigintiles from 1965 to the month $t-1$ to predict the next month t in columns IV and V. For the weights we use 0.05 (columns I and III) or 0.10 (columns II and IV) for the expected returns from equation (2) and 0.95 (columns I and III) and 0.90 (columns II and IV) for the bucket expected returns. The table shows the Sharpe ratios based on the realized (ex-post) monthly returns and standard deviation of these returns.

4.2 Impact of using different covariance matrices

Given the results in Tables II and III we fix the expected return to using the average in-sample return of the 20 buckets formed based on equation (2). In this section we will look at alternatives for the Fama and French 3-factor covariance matrix used in Tables II and III. In particular we show results for the sample covariance matrix measured over the past 24 to 60 months depending on the availability of the stocks, the Single Index Model (SIM) using only the market factor from the Fama and French factors, a shrunk version of the Fama and French 3-factor covariance matrix using equation (7) to shrink the factor loadings towards the cross-sectional mean, combinations of the sample covariance matrix and SIM, and combinations of the sample covariance matrix and the FF 3-factor covariance matrix. Table IV and V show the results.

Table IV: Optimization of the momentum strategy with different covariance matrices

	I	II	III	IV
Returns	20 buckets (feas)	20 buckets (feas)	20 buckets (feas)	20 buckets (feas)
Covariances	Sample Cov	SIM Cov	shrunk SIM Cov	shrunk F&F Cov
Constraints	None	None	None	None
$\lambda_1 = 100$	0.027	0.615	0.616	0.625
$\lambda_2 = 500$	0.021	0.615	0.616	0.625
$\lambda_3 = 1000$	0.111	0.614	0.616	0.626
$\lambda_4 = 5000$	-0.125	0.603	0.610	0.594
$\lambda_5 = 10000$	-0.153	0.584	0.587	0.570

Note: We solve the quadratic optimization problem in equation (1) for different choices for the risk-averseness parameter λ . For the expected return we use the average ex-post return of the momentum vigintiles from 1965 to the month $t-1$ to predict the next month t . For the covariance matrix we use the sample covariance matrix over the past 24 to 60 months (column I), the Single Index Model (equation (4) and (5) leaving out SMB and HML) with and without shrinkage of the market betas using equation (7), and the Fama and French 3-factor model in equations (4) and (5) with betas shrunk towards the cross-sectional mean, see equation (7). The table shows the Sharpe ratios based on the realized (ex-post) monthly returns and standard deviation of these returns.

The results in Table IV first of all show that the sample covariance matrix fails to produce any good results. Problem is that the number of assets easily exceeds the number of observations, resulting in the covariance matrix to be poorly defined. When placing more emphasis on risk (higher risk-averseness) we even get negative Sharpe ratios. Hence whereas the sample covariance matrix is a popular choice in the literature on robust optimization, it fails for problems with many assets.

Second the results in column II of Table IV show that the Single Index Model is a perfectly viable alternative for the 3-factor model (see Table II column V). We believe the reason is that the market beta dominates the long and short portfolio, and hence to reduce risk this is the most important dimension to match the long and short portfolio. Chan et al. (1999) find something similar in that for minimizing the variance of a long-only portfolio (all weights are non-negative) it does not matter when adding additional factors to the market factor.

Finally we see from columns III and IV in Table IV that shrinking the factor loadings to the cross-sectional mean only leads to a marginal improvement over simply using ordinary least squares to estimate the factor loadings. For example for the Single Index model we get very similar Sharpe ratios in Table IV (columns II & III) with and without shrinkage

The results in Table V focus on shrinkage as proposed in Ledoit and Wolf (2003). In particular we take the weighted average of the sample covariance matrix and the covariance matrix according to the Single Index Model. When comparing column I in Table V (99% weight for the Single Index Model, 1% weight for the sample covariance matrix) with column II in Table IV (100% weight for the Single Index Model covariance matrix) we see that shrinkage makes the results worse. Hence the quality of the sample covariance matrix is that low that combining it with the Single Index Model does not improve the Sharpe ratios. Giving more weight to the sample covariance matrix even leads to worse results. This contradicts the results of Ledoit and Wolf (2003). They use a better estimated sample covariance matrix, since it is based on fewer names (at most 1300) for a return history of 120 months, whereas we have 1500 to 2500 stocks with a return history of at most 60 months, but usually closer to 24 months. Both differences improve the performance of the sample covariance used by Ledoit and Wolf (2003). If we require our stocks to exist for at least 120 months the return to the standard momentum strategy would almost completely diminish. In addition we think that in this setting the beta dimension of risk for the momentum strategy is that important that measuring risk primarily through beta is the key to success. This also explains why the corresponding shrinkage results for the Fama and French 3-factor covariance matrix in columns III and IV of Table V do not produce higher Sharpe ratios.

Table V: Combined (sample and factor) covariance matrices

	I	II	III	IV
Returns	20 buckets (feas)	20 buckets (feas)	20 buckets (feas)	20 buckets (feas)
Covariances	0.99SIM+0.01Sample	0.95SIM+0.0Sample	0.99FF3+0.01Sample	0.95FF3+0.01Sample
Constraints	None	None	None	None
$\lambda_1 = 100$	0.600	0.539	0.597	0.528
$\lambda_2 = 500$	0.601	0.539	0.598	0.529
$\lambda_3 = 1000$	0.600	0.541	0.599	0.532
$\lambda_4 = 5000$	0.584	0.499	0.549	0.468
$\lambda_5 = 10000$	0.547	0.500	0.526	0.453

Note: We solve the quadratic optimization problem in equation (1) for different choices for the risk-averseness parameter λ . For the expected return we use the average ex-post return of the momentum vigintiles from 1965 to the month $t-1$ to predict the next month t . For the covariance matrix we use a combination of the sample covariance matrix over the past 24 to 60 months with either the Single Index Model (equation (4) and (5) leaving out SMB and HML) or the Fama and French 3-factor model in equations (4) and (5). The table shows the Sharpe ratios based on the realized (ex-post) monthly returns and standard deviation of these returns.

4.3 Impact of using weight constraints on individual stocks

Jaganathan and Ma (2003) show that to get the minimum variance portfolio it really helps to impose the constraint that weights are non-negative. Chan et al. (1999) even impose the weights to be bounded between 0 and 2%. This to some extent avoids error maximization because it avoids having extreme weights in individual stocks. These extreme weights are likely to be resulting from erroneous inputs.

Here we also test adding the constraint that the weights cannot exceed 1% in absolute value. Of course we already had the constraint that stocks that are ranked in the top half cannot have negative weights, and stocks that are ranked in the bottom half cannot have positive weights. The results are shown in Table VI.

First of all we look at the case where we use the raw returns in equation (2) as the return input, and the 3-factor covariance matrix. In Table II (column I) we have seen that especially when focusing on expected returns (very low risk-averseness parameter) we get a disappointing Sharpe ratio of 0.106. In that case the optimizer will really make use of the extreme expected returns resulting from equation (2). In Table VI (column III) we see that especially this case benefits from imposing a maximum weight of 1% on individual stocks. Hence the optimizer is no longer allowed to put almost all its weight in a few top ranked stocks. As a result the Sharpe ratio increases from 0.106 to 0.534. This, however, is still well short of Sharpe ratio of 0.617 when using expected returns based on 20 buckets (Table II, column V). Hence making the return input robust is better than imposing weight constraints. In fact, when doing so, adding the weight constraint does not help anymore as shown in Table VI (column IV) and comparing this result with column V in Table II. The robust return input already resulted in all weights being below 1% for individual stocks.

Second we see from columns I and II that also for the sample covariance matrix it helps to add the weight constraint when using the raw returns in equation (2) for the expected returns. And again (not reported here) it does not help adding the weight constraint when using the 20 buckets to determine the expected returns.

Table VI: Optimization of the momentum strategy with weight constraints

	I	II	III	IV
Returns	Raw returns	Raw returns	Raw returns	Vigintiles (feas)
Covariances	Sample Cov	Sample Cov	F&F Factor Cov	F&F Factor Cov
Constraints	None	1%	1%	1%
$\lambda_1 = 100$	-0.049	0.399	0.534	0.617
$\lambda_2 = 500$	-0.221	0.150	0.592	0.617
$\lambda_3 = 1000$	-0.186	0.097	0.596	0.617
$\lambda_4 = 5000$	-0.160	0.017	0.531	0.586
$\lambda_5 = 10000$	-0.191	0.027	0.475	0.566

Note: We solve the quadratic optimization problem in equation (1) for different choices for the risk-averseness parameter λ . For the expected return we use the raw returns from equation (2). For the covariance matrix we use the sample covariance matrix (columns I and II) or the 3-factor covariance matrix (column III). The table shows the Sharpe ratios based on the realized (ex-post) monthly returns and standard deviation of these returns.

5. Conclusion

In this paper we study quadratic optimization for a very large number of stocks. The danger in such a setting is that quadratic optimization turns into error maximization, with the optimizer trying to exploit especially the extreme expected returns and the extreme low risk elements in the covariance matrix. In the literature much attention is given to shrinkage estimators and weight constraints to reduce the impact of error maximization. This is often done in simple settings with poor return and covariance inputs, making it easy to improve the results.

The unique features in this paper are that we do use a useful input for both the expected returns and the covariance matrix, we have a large universe, and we have a hard-to-beat benchmark in the long-short equally weighted equity momentum portfolio. We find that in this case the Bayes-Stein estimator for expected returns, shrinkage of the factor loadings, shrinkage of the covariance matrix by combining different covariance matrices, and weight constraints all add little to nothing to improve the performance of the resulting portfolios. In fact these approaches are clearly worse than simply thinking carefully about logical robust expected returns for the specific problem and a covariance matrix that relates to the return source on the important risk dimensions. Hence we show in this practical case that common sense prevails over dogmatic theories that aim for a one-fits-all approach. In the process we also solved the initial non-convexity problems in aiming for a zero-investment long-short portfolio.

For further research it will be interesting to look into robust optimization. Challenge there is to refrain from the simplest way in which most studies apply this idea of taking into account the errors in the inputs during the optimization, as this basically leads to a standard quadratic optimization problem with more emphasis on risk. In addition it would be interesting to see whether robustification of the inputs and robust optimization can be used together to improve the results obtained by applying just one of the two approaches.

References

- Amihud, Y. (2002), Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31-56.
- Chan, L., J. Karceski and J. Lakonishok (1999), On portfolio optimization: Forecasting covariances and choosing the risk model, *Review of Financial Studies* 12, 937-974.
- DeMiguel, V., L. Garlappi, and R. Uppal (2007), Optimal versus naïve diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, forthcoming.
- Fama, E.F., and K.R. French (1993), Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.
- Fan, J., Y. Fan, and J. Lv (2008), High dimensional covariance matrix estimation using a factor model, *Journal of Econometrics* 147, 186-197.
- Goldfarb, D., and G. Iyengar (2003), Robust convex quadratically constrained programs, *Mathematical Programming* 97, 495-515.
- Jaganathan, R., and T. Ma (2003), Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651-1684.
- Jegadeesh, N., and S. Titman (1993), Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65-91.
- Jegadeesh, N. and S. Titman (1995), Overreaction, delayed reaction, and contrarian profits, *The Review of Financial Studies* 8, 973-993
- Jorion, P. (1986), Bayes-Stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21, 279-292.
- Ledoit, O., and M. Wolf (2003), Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* 10, 603-621.
- Ledoit, O., and M. Wolf (2004), “Honey, I shrunk the sample covariance matrix”, *Journal of Portfolio Management* 30, 110-119
- Lehmann, B. (1990), Fads, martingales, and market efficiency, *Quarterly Journal of Economics* 105, 1-28.
- Lo, A. and A.C. MacKinlay (1990), When are contrarian profits due to stock market overreaction?, *Review of Financial Studies* 3, 175-205.
- Martens, M., and A. van Oord (2008), Hedging the time-varying exposures of momentum returns. Available at SSRN: <http://ssrn.com/abstract=1099932>.
- Michaud, R. (1998), *Efficient asset management: A practical guide to stock portfolio optimization*, Oxford University Press.
- Pastor, L., and R.F. Stambaugh (2003), Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.
- Stein, C. (1955), Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proceedings of the 3rd Berkeley Symposium on Probability and Statistics I*, 197-206.