Robust optimization of time series momentum portfolios

Jeremy Fague†
Caio Almeida‡

Abstract Mean-Variance Optimization (MVO) is well-known to be extremely sensitive to slight differences in the expected returns and covariances: if these measures change day to day, MVO can specify very different portfolios. Making wholesale changes in portfolio composition can cause the incremental gains to be negated by trading costs. We present a method for regularizing portfolio turnover by using the $\ell_1$ penalty, with the amount of penalization informed by recent historical data. We find that this method dramatically reduces turnover, while preserving the efficiency of mean-variance optimization in terms of risk-adjusted return. Factoring in reasonable estimates of transaction costs, the turnover-regularized MVO portfolio substantially outperforms a leverage-constrained MVO approach, in terms of risk-adjusted return.

Keywords: Mean Variance; Momentum; Turnover; Regularization.
JEL Code: C51, G11, G12.

1. Introduction

For as long as investors have sought capital appreciation by purchasing equities, they have developed strategies to capture returns, thereby compensating them for accepting different sources of risk. No matter the strategy, investors seek an optimal method for determining how to allocate their capital among various investment choices.

In his seminal paper, Markowitz (1972) describes the process of utilizing information about the expected returns, variances, and covariances of financial assets to construct a portfolio, making optimal trade-offs between risk and return. With $n$ assets, the basic problem of portfolio optimization is

$$\min_{w \in \mathbb{R}^n} w^T \Sigma w$$

s.t.

$$\alpha^T w \geq r_E$$

$$\sum_{i=1}^{n} w_i = 1$$
where $\alpha \in \mathbb{R}^n$ is a vector of expected returns, $\Sigma \in \mathbb{R}^{n \times n}$ is the covariance matrix of assets, $w$ are the portfolio weights, and $r_E$ is a target expected return.

While the approach is sensible from a theoretical perspective, in practice the optimization is extremely sensitive to the inputs. Slight changes in the expected returns $\alpha$ or the covariance matrix $\Sigma$ can result in drastically different portfolios. Thus any investor, using mean-variance optimization to decide allocations to a large number of assets, will likely experience substantial instability in the optimal solution to the Markowitz problem. Further, if the first and second moments of the data differ in the future from what they have been in the past, the parameters are misspecified. The resulting portfolio may have quite different realized returns and volatility than expected. This issue calls for the application of methods to make the solution robust to small changes in the inputs. In particular, regularization provides interesting possibilities. In this paper, we provide a new methodology based on an $\ell_1$ regularization of portfolio weights towards the weights obtained on a previous day. We show that this method substantially reduces turnover and obtains high out-of-sample Sharpe ratios, even in the presence of transaction costs.

The remainder of the paper is organized as follows. In Section 2 we discuss the related literature. In Section 3, we describe the all the data adopted, while in Section 4 we introduce our portfolio regularization method. Section 5 presents the empirical results, comparing our shrinkage estimator to the traditional mean-variance estimator with short-sale constraints suggested by Brodie et al. (2009). Section 6 offers conclusions and potential topics for future research.

2. Literature Review

Black and Litterman (1992) develop one of the earliest methods for stabilizing mean-variance optimization. Their approach relies on CAPM, determining the equilibrium expected returns based on market-clearing holdings of all market participants, and allowing individual investors to add their own views to the model to construct a “tilt” away from the market portfolio. This model effectively shrinks the expected returns used in mean-variance optimization toward a market-based equilibrium view, generating robustness to investors’ specifications of expected returns.

As regularization methods have become ubiquitous in statistics, machine learning, and other applications, a rich body of work on the benefits of applying regularization methods to mean-variance optimization has developed. Jagannathan and Ma (2003) demonstrate that imposing a nonnegativity constraint on the portfolio weights (essentially constraining the 1-norm of the portfolio weights to be 1) results in comparable performance to other regu-
larization methods. Similarly, Brodie et al. (2009) impose a 1-norm penalty on the portfolio weights in the objective function — encouraging sparsity and allowing for negative weights, but not large weights — and find that the resulting portfolios outperform the equally-weighted portfolio in terms of the Sharpe ratio. DeMiguel et al. (2009) examine the effects of imposing a number of different regularizations of the portfolio weights, including the 1-norm, 2-norm, and covariance-matrix norm, and find that the norm-constrained portfolios often have impressive Sharpe ratios compared to benchmarks, including the portfolio formed using the methodology of Jagannathan and Ma (2003).

Most similar to our method is the regularization by Kourtis (2015), in which the portfolio return constraint is \( \alpha^T w - \kappa \| w - w_0 \|_1 = r_E \), where \( \kappa \) is the transaction cost and \( w_0 \) are the portfolio weights before rebalancing. With a 1% transaction cost, turnover levels were leveled with those of an equal-weighted portfolio; however, in terms of the Sharpe ratio, this regularized portfolio was unable to consistently beat the equal-weighted portfolio.

The key difference between the method suggested in this paper and that of Kourtis (2015) is the treatment of the turnover regularization. Kourtis (2015) incorporates transaction costs directly into the expected-return constraint so that any trading which does not bring the after-transaction-cost return below the constraint is allowed. Thus he explicitly informs the maximum amount of turnover that is allowed each day by considering the economics of trading. In contrast, the method of this paper does not directly constrain transaction costs to be below a certain level. Turnover is penalized in the objective function, with the magnitude of penalization informed by the data via cross-validation. The range of potential values for the parameter that controls penalization (\( \lambda \)) includes levels which are similar to transaction costs for institutional asset managers, but also levels which are significantly higher. Including such a range of candidate values for \( \lambda \) allows for control of overfitting via cross-validation. Thus it is not only economic concerns that factor into the selection of \( \lambda \), but also statistical concerns: the model seeks to both control transaction costs and introduce robustness to constantly-changing or misspecified data.

Other literature on time series momentum and transaction costs also informs the methodology presented in this paper. Moskowitz et al. (2012) document the time series momentum asset pricing anomaly in futures and forwards across a wide array of equity indices, FX, commodities, and government bonds: the return of these instruments over the last 12 months is positively correlated to future return. D’Souza et al. (1998) find that this phenomenon is also found in equity returns and has persisted for almost a full century. Frazzini et al. (2018) conduct an in-depth study of actual trading costs for a large institutional manager and find them to be much lower than
previously estimated in academia. Particularly relevant for this paper, they estimate trading costs for a long-short portfolio in large-cap stocks at about 15 basis points.

3. Data: Expected Returns and Covariances for S&P 500 Stocks

We gather data on daily returns and index composition of S&P 500 stocks from 1926 to 2019, from Wharton Research Data Services. We calculate daily expected returns and covariances as follows:

- Based on historical gross returns \( \{r_{ui}\}_{t-252}^{t-1} \) for each firm \( i \) in the index on day \( t \), calculate the geometric average using gross returns from day \( t - 252 \) to day \( t - 1 \) and subtract one to make a net return. These became the TSMOM signals on day \( t \). That is, for each \( i = 1, \ldots, N_t \), where \( N_t \) is the number of firms in the index on day \( t \), define \( TSMOM_{ti} = \left( \prod_{u=t-252}^{t-1} r_{ui} \right)^{\frac{1}{252}} - 1 \).

- Calculate the historical covariance matrix of returns from \( t - 252 \) to \( t - 1 \).

What is today called the S&P 500 stock index did not, in fact, always have 500 stocks, and in recent years it has had more than 500 stocks in order to include a few companies with dual share classes.\(^1\) When the index was first published in 1926, it had the name S&P 90 since it had 90 stocks; it was expanded to 500 in early 1957. Since the index has always represented a diverse cross-section of American economic activity, this paper uses it in its multiple forms over the years as a selection of relatively large, liquid stocks that should be easily tradable.

When data are missing for a particular stock on day \( t \) — in particular, if any data within the previous 252 days are missing for that stock — that stock is excluded from the portfolio optimization on day \( t \). This helps ensure that the resulting historical covariance matrices are positive definite: if covariances is permitted to be calculated with only the available days, the resulting matrices are not always positive definite, rendering the optimization problem ill-conditioned. Since the number of stocks to choose from is high, the size of the optimization problem remains large over time, despite having to periodically drop some stocks due to missing data.

\(^1\)Standard & Poor’s has instituted a rule banning companies with dual share classes from being included in the index; the current constituents with dual share classes were grandfathered in.
4. Turnover Regularization

To create a benchmark MVO portfolio through time, the following optimization problem is solved daily:

\[
\min_{w \in \mathbb{R}^n} \quad w^T \Sigma_t w \\
\text{s.t.} \quad \alpha_t^T w \geq r_E \\
\sum_{i=1}^{n} w_i = 1 \\
\|w\|_1 \leq 1.5.
\]

The last constraint limits shorting and absolute position size while encouraging sparsity, creating more stable portfolios through time, as described by Brodie et al. (2009). Without this constraint, MVO results in wildly unstable portfolios through time, making it a subpar benchmark. We denominate this model the leverage-constrained MVO.

The method suggested in this paper consists of solving the following optimization problem daily:

\[
\min_{w \in \mathbb{R}^n} \quad w^T \Sigma_t w + \lambda \|w - w_{t-1}^*\|_1 \\
\text{s.t.} \quad \alpha_t^T w \geq r_E \\
\sum_{i=1}^{n} w_i = 1 \\
\|w\|_1 \leq 1.5.
\]

The additional $\ell_1$ penalty in the objective function shrinks the solution $w$ towards the currently-held portfolio $w_{t-1}^*$. These prior weights $w_{t-1}^*$ include the effects of the prior day’s returns, so that they are in fact the “currently held” portfolio before any rebalancing today. Naturally, this reliance on prior portfolio weights necessitates an initial portfolio: $w_0$ is simply the benchmark MVO portfolio from above.

The parameter $\lambda$ is fitted via cross-validation: the “best” value is selected from a set of 10 possibilities. The historical data are essentially split into “blocks” of 252 days each. The first block serves as training data for the second block; then the second block serves as training data for the third block, and so on. To tune $\lambda$, the model is run over the training period for each candidate value, and the value leading to the highest portfolio return is selected. This value is then used to run the model over the next block. Thus in subsequent blocks, the training method essentially tunes $\lambda$ to what would
have been optimal over the training block, then uses that $\lambda$ for the subsequent block.

Optimization problems were solved using CVXPY, the Python implementation of CVX, a package for solving disciplined convex programs. For several dates, the solver failed to find a solution. In these cases, the weights from the previous day augmented or diminished by their returns over the previous day ($w_{t-1}^*$ defined above) were used. On other occasions, the list of stocks considered for inclusion in the portfolio changed due to either data availability or changes in the index. In these cases, positions in any stocks removed from the index were fully liquidated.

5. Benchmark versus Turnover Regularized MVO

5.1 Leverage-constrained Mean-Variance Optimization Portfolio

The cumulative returns of the benchmark MVO portfolio (leverage-constrained) are shown in Figures A1 (absolute scale) and A2 (log scale) which allows us to better compare the performance of the two estimators from 1926 to 1980. Overall summary statistics are reported in Table A1. At first glance, it performs quite well over time. It is consistently profitable, especially after the market crash and subsequent drawdown leading into the Great Depression. Its average annual return (calculated as a geometric average) over the entire period is 6.76%, and its average annual volatility (calculated as $\sqrt{252}$ times the daily standard deviation of returns) is 9.34%. The average annual Sharpe ratio over the entire period, calculated by averaging the rolling 252-day Sharpe ratio, is 0.984. As reported in Table A2, the risk-adjusted return is persistent over decades: only in the 1920s, 1960s, 1970s, and 1990s was the Sharpe ratio below 1, and among those four decades the lowest average Sharpe ratio was 0.965, in the 1970s.

The 1-norm constraint on the weights effectively allows a maximum short position of 25% of capital; thus the largest possible long position is 125% of capital. With the exception of one day in 1932, the benchmark MVO portfolio maximizes the use of short sales every day, as shown in the time-series plot of total exposure in Figure A3. Additionally, while the 1-norm constraint would theoretically lead to some sparsity in the portfolios, this is not generally the case for the benchmark MVO portfolio. Figures A4 and A5 show, respectively, the number of stocks in the benchmark MVO portfolio through time and the percentage of available stocks used through time. Figures A6 and A7 present these measurements in histograms to illustrate the distribu-

---

2Sharpe ratios in this paper are calculated as the cumulative excess return over 252 days, divided by the annualized standard deviation of returns over that period.
tion over the testing period. On 95% of days, it uses all stocks in the S&P 500; on the other days, it generally uses more than 90%. It appears that the benefits to maintaining positions in most stocks, even if the total portfolio is 125/25 long/short, outweighs any benefits of sparsity in the benchmark MVO portfolio.

The major weakness of the benchmark MVO portfolio is in its turnover. On average, turnover is 15% of capital every day, or 38 times invested capital every year, as reported in Table A1. These numbers are enormous: transaction costs have a debilitating effect on returns. Using the estimates of Frazzini et al. (2018), trading costs in large-cap stocks for large institutional investment managers amount to approximately 15 basis points of the amount traded. To determine the effect of transaction costs on the benchmark MVO portfolio, after-transaction-cost (ac) returns are calculated as

$$r_{ac,t} = r_t - 0.0015 \| w_t - w_t^{*} \|_1$$

and the previous results are recomputed.

Figures A8 and A9 show the cumulative return over time in absolute and log scales, respectively, net of transaction costs, of the benchmark and reg-turnover MVO portfolios, whereas Table A1 reports overall summary statistics. The average annual return of the benchmark MVO portfolio falls to just 0.829%, while its average annual volatility slightly increases to 9.35%. The average annual Sharpe ratio dives to 0.137, demonstrating that inefficiency in turnover translates directly into inefficiency in risk-adjusted returns. The risk-adjusted return after transaction costs is also persistent across decades, as presented in Table A3. In recent years returns improve, but no decade has an after-transaction-cost average Sharpe ratio higher than 0.5.

### 5.2 Regularized Turnover Mean-Variance Optimization Portfolio

Figures A1 and A2 show, in absolute and log scales, respectively, the cumulative returns of the mean-variance optimization portfolio with an additional $\ell_1$ constraint on turnover (hereafter reg-turnover in short) alongside those of the benchmark MVO portfolio. Table A1 reports overall summary statistics. The reg-turnover MVO portfolio outperforms the benchmark MVO portfolio even before taking transaction costs into account. Its average annual return over the entire period is 9.75%, and its average annual volatility is 10.69%. The average annual Sharpe ratio over the entire period, calculated by averaging the rolling 252-day Sharpe ratio, is 1.156. As reported in Table A2, the high Sharpe ratio persists across decades: only in the 1990s is the Sharpe ratio below 1, and even then it is close.
Unlike the benchmark, the reg-turnover MVO portfolio often uses noticeably less than the full 125% long/25% short total exposure — on 29% of all days, in fact. The total exposure through time is shown alongside that of the benchmark MVO portfolio in Figure A3. The differences are usually small — on most days with total positions less than the full 150% of capital, the total positions are at least 148% — but there are hundreds of days when the total positions are less than that, going as low as 140% or less on 122 days. While these differences are somewhat small, they suggest that the regularization of turnover also acts as an additional penalty on total exposure beyond the benchmark MVO constraints.

Sparsity is more prevalent in the reg-turnover MVO portfolio than in the benchmark. Figures A4 and A5 show, respectively, the number of stocks in the reg-turnover MVO portfolio and the percentage of available stocks used through time, alongside the same measurements for the benchmark MVO portfolio. Figures A6 and A7 present these measurements in histograms to illustrate the distribution over the testing period. On 18% of days, the reg-turnover MVO portfolio uses fewer than 95% of the stocks in the S&P 500, and on more than half of those days, the percentage is 80% or below. Thus it appears that the regularization of turnover increases portfolio sparsity beyond the levels imposed by the exposure constraint.

Perhaps the most exciting feature of the reg-turnover MVO portfolio is its limited turnover, as reported in Table A1. On average, turnover is just 0.58% of capital each day, or 1.45 times invested capital every year. These numbers are significantly smaller than the benchmarks of the benchmark MVO portfolio. To determine the effect of transaction costs on the reg-turnover MVO portfolio, after-transaction-cost returns are calculated in the same manner as above for the Benchmark MVO portfolio, and the previous results are recomputed.

The after-transaction-cost average annual return of the reg-turnover MVO portfolio falls slightly to 9.512%, while its average annual volatility stays the same at 10.69%. The average annual Sharpe ratio remains high at 1.132, demonstrating that the efficiency in managing turnover mostly preserves risk-adjusted returns. The risk-adjusted returns after transaction costs are also persistent across decades, as presented in Table A3: comparing the average annual Sharpe ratios after transaction costs to those before transaction costs in each decade, the numbers are only slightly lower after transaction costs.

5.3 Statistical Testing

We conduct hypothesis testing to determine whether turnover regularization has statistically significant effects on portfolio returns and Sharpe ratios,
as well as whether the penalty term leads to differences in these measures. Table A4 presents the results of the latter exercise.

In order to perform the test described above, for each value of the turnover penalty term $\lambda$, we calculate daily portfolio returns and excess returns, divide excess returns by the standard deviation of the portfolio returns, and then conduct $t$-tests for differences in means. For most values of the penalty parameter, the mean Sharpe ratio of the reg-turnover MVO portfolio exceeds that of the benchmark MVO portfolio, after transaction costs, at confidence levels ranging from 90\% to 99.9\%. However, the outperformance in risk-adjusted returns of the reg-turnover MVO portfolio is not robust to the magnitude of the penalty term applied: for two values of $\lambda$ (0.03125 and 1), the $p$-values are higher than 0.3, indicating that we cannot really statistically distinguish between the Sharpe ratio of the reg-turnover and that of the benchmark MVO.

Thus, the outperformance of the reg-turnover MVO portfolio in terms of risk-adjusted returns cannot be assumed to hold true for any level of turnover regularization. It is certainly possible that limitations in the data are to blame for the insignificance of the results for some levels of regularization. For example, the parameter $\lambda$ is only selected to be 1.0 once, so there is only one 252-day period that uses it. On the other hand, we report one strongly statistically significant case when the level of regularization is the highest tested ($\lambda = 4$), that clearly distinguishes the performance of the two portfolio estimators. In this case, the reg-turnover MVO portfolio has a significantly better Sharpe ratio than the benchmark MVO portfolio, with a $p$-value much lower than 0.1\%.

6. Conclusions and Further Research Opportunities

In this paper we compare a standard leverage-constrained method of implementing mean-variance optimization (resulting in the “benchmark MVO” portfolio) to another method that regularizes portfolio turnover by adding an $\ell_1$ penalty to discourage daily turnover (resulting in the “reg-turnover MVO” portfolio). The results show that even before transaction costs, the risk-adjusted returns of the reg-turnover MVO portfolio are higher than those of the benchmark MVO portfolio. Factoring in transaction costs, the gulf between the two portfolios widens significantly, and furthermore, the reg-turnover MVO portfolio loses little of its risk-adjusted return to these costs. The resulting difference in risk-adjusted returns is statistically significant at the 0.1\% level for the highest level of regularization used in the experiment, lending additional credence to the idea that regularizing turnover results in portfolios that are efficient, both in terms of risk-adjusted return and transaction costs.
Further research could be done to examine the effects of some of the implementation details of this experiment. This method selects the regularization parameter $\lambda$ by maximizing return over a historical sample; other selection methods based on other metrics (minimizing variance, maximizing Sharpe ratio, etc.) or changing other parameters (the number of historical days used to fit the parameter, the range of candidate values, using an information criterion rather than cross-validation, etc.) could lead to different results. Additionally, transaction costs are not necessarily constant across all stocks, even if all are large-cap names: liquidity might vary among the candidate stocks. A version of the method presented here that uses a “group LASSO” approach, to first group stocks by liquidity and then penalize each group according to its relative liquidity, could potentially decrease transaction costs further.

References


A. Figures and Tables

Figure A1
Cumulative portfolio returns, before transaction costs

Figure A2
Cumulative portfolio returns, before transaction costs - log scale
Figure A3
Total portfolio exposure

Figure A4
Number of assets used
Figure A5
Percent of assets used

Figure A6
Number of assets used - histogram
Robust optimization of time series momentum portfolios

Figure A7
Percent of assets used - histogram

Figure A8
Cumulative portfolio returns, after transaction costs
Figure A9
Cumulative portfolio returns, after transaction costs - log scale

Figure A10
Rolling annual Sharpe ratio, after transaction costs
Figure A11
Shrinkage Parameter

Table A1
Summary portfolio statistics (annualized, except drawdown)

<table>
<thead>
<tr>
<th></th>
<th>before t-costs</th>
<th>after t-costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>benchmark</td>
<td>reg-turnover</td>
</tr>
<tr>
<td>return</td>
<td>6.76%</td>
<td>9.75%</td>
</tr>
<tr>
<td>volatility</td>
<td>9.34%</td>
<td>10.69%</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>0.984</td>
<td>1.156</td>
</tr>
<tr>
<td>turnover</td>
<td>3800%</td>
<td>145%</td>
</tr>
<tr>
<td>max drawdown</td>
<td>82.87%</td>
<td>71.99%</td>
</tr>
</tbody>
</table>
### Table A2

Sharpe ratios by decade, before transaction costs

<table>
<thead>
<tr>
<th>Decade</th>
<th>Benchmark</th>
<th>Reg-turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920s</td>
<td>0.9837</td>
<td>1.1560</td>
</tr>
<tr>
<td>1930s</td>
<td>1.0001</td>
<td>1.1707</td>
</tr>
<tr>
<td>1940s</td>
<td>1.1506</td>
<td>1.2958</td>
</tr>
<tr>
<td>1950s</td>
<td>1.2036</td>
<td>1.3424</td>
</tr>
<tr>
<td>1960s</td>
<td>0.9836</td>
<td>1.1752</td>
</tr>
<tr>
<td>1970s</td>
<td>0.9655</td>
<td>1.1579</td>
</tr>
<tr>
<td>1980s</td>
<td>1.1354</td>
<td>1.2352</td>
</tr>
<tr>
<td>1990s</td>
<td>0.9818</td>
<td>0.9962</td>
</tr>
<tr>
<td>2000s</td>
<td>1.0959</td>
<td>1.0942</td>
</tr>
<tr>
<td>2010s</td>
<td>1.2415</td>
<td>1.3410</td>
</tr>
</tbody>
</table>

### Table A3

Sharpe ratios by decade, after transaction costs

<table>
<thead>
<tr>
<th>Decade</th>
<th>Benchmark</th>
<th>Reg-turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920s</td>
<td>0.1368</td>
<td>1.1320</td>
</tr>
<tr>
<td>1930s</td>
<td>0.1390</td>
<td>1.1465</td>
</tr>
<tr>
<td>1940s</td>
<td>0.2070</td>
<td>1.2731</td>
</tr>
<tr>
<td>1950s</td>
<td>0.1960</td>
<td>1.3214</td>
</tr>
<tr>
<td>1960s</td>
<td>−0.0129</td>
<td>1.1552</td>
</tr>
<tr>
<td>1970s</td>
<td>0.0931</td>
<td>1.1387</td>
</tr>
<tr>
<td>1980s</td>
<td>0.3233</td>
<td>1.2169</td>
</tr>
<tr>
<td>1990s</td>
<td>0.1862</td>
<td>0.9783</td>
</tr>
<tr>
<td>2000s</td>
<td>0.3448</td>
<td>1.0742</td>
</tr>
<tr>
<td>2010s</td>
<td>0.4843</td>
<td>1.3234</td>
</tr>
</tbody>
</table>
Table A4
Testing difference in means, after transaction costs

<table>
<thead>
<tr>
<th>regularization parameter ($\lambda$)</th>
<th>$p$-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.007812</td>
<td>0.058542</td>
</tr>
<tr>
<td>0.015625</td>
<td>0.083860</td>
</tr>
<tr>
<td>0.031250</td>
<td>0.322408</td>
</tr>
<tr>
<td>0.062500</td>
<td>0.065351</td>
</tr>
<tr>
<td>0.125000</td>
<td>0.096647</td>
</tr>
<tr>
<td>0.250000</td>
<td>0.056743</td>
</tr>
<tr>
<td>0.500000</td>
<td>0.017986</td>
</tr>
<tr>
<td>1.000000</td>
<td>0.378055</td>
</tr>
<tr>
<td>2.000000</td>
<td>0.071659</td>
</tr>
<tr>
<td>4.000000</td>
<td>0.000002</td>
</tr>
</tbody>
</table>