

Mean-Reversion and Optimization

Zura Kakushadze^{§†‡1}

[§] *Quantigic[®] Solutions LLC*

*1127 High Ridge Road #135, Stamford, CT 06905*²

[†] *Department of Physics, University of Connecticut*

1 University Place, Stamford, CT 06901

[‡] *Free University of Tbilisi, Business School & School of Physics*
240, David Agmashenebeli Alley, Tbilisi, 0159, Georgia

(August 9, 2014; revised September 22, 2014)

Abstract

The purpose of these notes is to provide a systematic quantitative framework – in what is intended to be a “pedagogical” fashion – for discussing mean-reversion and optimization. We start with pair trading and add complexity by following the sequence “mean-reversion via demeaning → regression → weighted regression → (constrained) optimization → factor models”. We discuss in detail how to do mean-reversion based on this approach, including common pitfalls encountered in practical applications, such as the difference between maximizing the Sharpe ratio and minimizing an objective function when trading costs are included. We also discuss explicit algorithms for optimization with linear costs, constraints and bounds. We also illustrate our discussion on an explicit intraday mean-reversion alpha.

¹ Email: zura@quantigic.com

² DISCLAIMER: This address is used by the corresponding author for no purpose other than to indicate his professional affiliation as is customary in publications. In particular, the contents of this paper are not intended as an investment, legal, tax or any other such advice, and in no way represent views of Quantigic Solutions LLC, the website www.quantigic.com or any of their other affiliates.

1 Introduction and Summary

Statistical Arbitrage (StatArb) “refers to highly technical short-term mean-reversion strategies involving large numbers of securities (hundreds to thousands, depending on the amount of risk capital), very short holding periods (measured in days to seconds), and substantial computational, trading, and information technology (IT) infrastructure” (Lo, 2010). So, what is this “mean-reversion”?³ The basic idea is simple: some quantities are historically correlated, sometimes these correlations are temporarily undone by some unusual market conditions, but one expects – or rather hopes – that the correlation will be restored in the future. StatArb tries to capture a profit from such temporary mispricings.

The purpose of these notes is to provide a systematic quantitative framework – in what is intended to be a “pedagogical” fashion – for discussing mean-reversion and optimization. There are a myriad ways of doing (*i.e.*, implementing) mean-reversion. One such approach can be schematically described via a sequence “mean-reversion via demeaning → regression → weighted regression → (constrained) optimization → factor models”. These notes follow precisely this sequence, starting from the most basic form of StatArb, pair trading, and gradually adding complexity. This naturally introduces mean-reversion around means of returns, regression and, ultimately, optimization and factor models – via the observation that weighted regression is nothing but a zero specific risk⁴ limit of optimization with a factor model.

Within this framework we discuss various important intricacies and pitfalls that arise in practical applications, often overlooked, deemphasized and/or not addressed, including in commercially available offerings. Should regression weights used in conjunction with optimization be based on historical or specific risk? How should one optimize regressed returns? How does one include constraints into optimization? Is optimization based on objective function minimization the same as maximizing the Sharpe ratio once costs are included? How does one optimize with linear costs, constraints and bounds? *Etc.* These are some of the topics we discuss in these notes – systematically and “pedagogically”, we hope.

The organization of these notes is as follows. Section 2 discusses mean reversion: pair trading → multiple stocks → multiple binary clusters (industries) → regression → non-binary generalization → weighted regression. Section 3 discusses optimization: maximizing Sharpe ratio → adding multiple linear constraints (including dollar neutrality) → regression as a limit of optimization → factor models → optimization with a factor model with linear constraints (including pitfalls). Section 4 is an “intermezzo” of sorts, which is kept on the lighter side, to help digest Sections 2 and 3 before adding even more complexity in Sections 5 and 6. Section 5 discusses optimization with constraints and costs, including the difference between Sharpe

³ “Mean-reversion strategy” is mostly trader lingo – which is what the author is accustomed to. Academic finance literature mostly uses “contrarian investment strategy” instead. This paper uses the term “mean-reversion (strategy)” throughout.

⁴ “Specific risk” is multi-factor risk model terminology. Some may prefer “idiosyncratic risk”.

ratio maximization and minimizing an objective function, and when the latter can be used as an approximation for the former. Section 6 discusses explicit algorithms for optimization with linear costs, constraints and bounds, including in the context of factor models. Section 7 illustrates the regression approach discussed in Section 2 by giving an explicit example of an intraday mean-reversion alpha (with a 5-year simulated performance), based on overnight returns and an industry classification, together with additional bells and whistles for risk management and dealing with outliers. Section 8 contains brief concluding remarks.

2 Mean-Reversion

2.1 Pair Trading

Often, when explaining StatArb popularly, a pair trading example is given. In a nutshell, it goes as follows. Suppose you have two historically correlated stocks in the same sector, stock A and stock B (*e.g.*, Exxon Mobil (XOM) and Royal Dutch Shell (RDS.A)). If, temporarily, stock A moves up (A is rich) while stock B moves down (B is cheap), the pair trading strategy amounts to shorting A and buying B in such proportion that the total position is dollar neutral. Dollar neutrality ensures that the position is (approximately) insensitive to overall market movements – it is simply a hedge against market risk.

Intuitively, this all makes sense assuming the spread between A and B converges back to its historical values. This is the essence of mean-reversion. The money is made from a temporary mispricing in A and B. However, upon a second look, one may ask: How do I know if A is rich and B is cheap? Indeed, first, A and B typically trade at different prices to begin with. Second, the prices of A and B are not constant on average – typically, albeit not always, they each have an upward drift. So, how does one quantify “rich” and “cheap” in pair trading?

2.2 Returns, Not Prices

It is not prices but returns that define “rich” and “cheap”. The idea is that on average stocks A and B are expected to move in sync. Say both move up. If A moves up more than B on a relative basis to their respective prices, then A is rich and B is cheap. Let $P_A(t_1)$ and $P_B(t_1)$ be the prices of A and B at time t_1 , and let $P_A(t_2)$ and $P_B(t_2)$ be the prices of A and B at a later time t_2 . (*E.g.*, t_1 can be yesterday’s close – with $P_A(t_1)$ and $P_B(t_1)$ adjusted for any splits and dividends if the ex-date is today – and t_2 can be today’s open.) The corresponding returns are

$$R_A = \frac{P_A(t_2)}{P_A(t_1)} - 1 \tag{1}$$

$$R_B = \frac{P_B(t_2)}{P_B(t_1)} - 1 \tag{2}$$

Since typically these returns are small, we can use an alternative definition:

$$R_A \equiv \ln \left(\frac{P_A(t_2)}{P_A(t_1)} \right) \quad (3)$$

$$R_B \equiv \ln \left(\frac{P_B(t_2)}{P_B(t_1)} \right) \quad (4)$$

So, the mean-reversion idea in pair trading can now be quantified as follows. If $R_A > R_B$, then A is rich, B is cheap, short A and buy B.

We can conveniently restate this using the demeaned returns \tilde{R}_A and \tilde{R}_B :

$$\bar{R} \equiv \frac{1}{2} (R_A + R_B) \quad (5)$$

$$\tilde{R}_A \equiv R_A - \bar{R} \quad (6)$$

$$\tilde{R}_B \equiv R_B - \bar{R} \quad (7)$$

where \bar{R} is the mean return.⁵ Now a stock is rich if its demeaned return is positive, and it is cheap if its demeaned return is negative. So, assuming the returns have been demeaned, we short positive return stocks and buy negative return stocks.

In the case of 2 stocks, the numbers of shares Q_i , $i = A, B$ to short/buy are fixed by the total desired dollar investment I and the requirement of dollar neutrality:

$$P_A |Q_A| + P_B |Q_B| = I \quad (8)$$

$$P_A Q_A + P_B Q_B = 0 \quad (9)$$

where P_i are the prices at the time the position is established, $Q_i < 0$ for short-sales, and $Q_i > 0$ for buys. Here we assume no leverage and 0 margins (see footnote 6).

2.3 Generalization to Multiple Stocks

What if we have more than two historically correlated stocks in the same sector? (*e.g.*, Exxon Mobil, Royal Dutch Shell, Total (TOT), Chevron (CVX) and BP (BP)). While we can do pair trading for each pair of stocks from such a set, can we have a mean-reversion strategy for the entire set? Demeaned returns make this a breeze.

Let R_i , $i = 1, \dots, N$ be the returns for our historically correlated N stocks:

$$R_i = \ln \left(\frac{P_i(t_2)}{P_i(t_1)} \right) \quad (10)$$

$$\bar{R} \equiv \frac{1}{N} \sum_{i=1}^N R_i \quad (11)$$

$$\tilde{R}_i \equiv R_i - \bar{R} \quad (12)$$

⁵ Here and in the following \bar{R} refers to the *cross-sectional* mean return (not the time series mean return). Also, \tilde{R}_A , \tilde{R}_B and \tilde{R}_i (see below) refer to the deviation from the mean return \bar{R} .

So, following our intuition from the 2-stock example, we can short stocks with positive \tilde{R}_i and buy stocks with negative \tilde{R}_i . We have the following conditions:⁶

$$\sum_{i=1}^N P_i |Q_i| = I \quad (13)$$

$$\sum_{i=1}^N P_i Q_i = 0 \quad (14)$$

Here: I is the total desired dollar investment; (14) is dollar neutrality; $Q_i < 0$ for short-sales; $Q_i > 0$ for buys; P_i are the prices at the time the position is established. We have 2 equations and $N > 2$ unknowns. So, we need to specify how to fix Q_i .

A simple way of specifying Q_i is to have the dollar positions

$$D_i \equiv P_i Q_i \quad (15)$$

proportional to the demeaned returns:

$$D_i = -\gamma \tilde{R}_i \quad (16)$$

where $\gamma > 0$ (recall that we short $\tilde{R}_i > 0$ stocks and buy $\tilde{R}_i < 0$ stocks). Then (14) is automatically satisfied as $\sum_{i=1}^N \tilde{R}_i = 0$ by definition, while (13) fixes γ :

$$\gamma = \frac{I}{\sum_{i=1}^N |\tilde{R}_i|} \quad (17)$$

Eq. (16) defines *one* mean-reversion strategy. There are a myriad of them. One drawback of (16) is that, by construction, on average it will take larger positions in more volatile stocks (as volatile stocks on average have larger $|\tilde{R}_i|$). Below we will discuss risk management and other ways of constructing D_i , *i.e.*, other mean-reversion strategies. Let us discuss a further generalization first.

2.4 Generalization to Multiple Clusters

We will refer to each group of stocks for which we can perform the analysis of the previous subsection as “clusters”. Depending on a given industry classification scheme, such clusters are called different names, such as industries, sub-industries, *etc.*⁷ Let there be K clusters labeled by $A = 1, \dots, K$. Let Λ_{iA} be an $N \times K$ matrix such that if the stock labeled by i ($i = 1, \dots, N$) belongs to the cluster labeled by

⁶ We assume no leverage and 0 margins. Nontrivial leverage simply rescales the investment level I . If margins are present, on top of I invested in stocks, we need an additional amount I' to maintain margins, which simply reduces the strategy return due to the borrowing interest rate.

⁷ *E.g.*, we could have one group of stocks from the oil sector, the second group from technology, and the third group from, say, healthcare.

A , then $\Lambda_{iA} = 1$; otherwise, $\Lambda_{iA} = 0$. We will assume that each and every stock belongs to one and only one cluster (so there are no empty clusters), *i.e.*,

$$N_A \equiv \sum_{i=1}^N \Lambda_{iA} > 0 \quad (18)$$

$$N = \sum_{A=1}^K N_A \quad (19)$$

We have

$$\Lambda_{iA} = \delta_{G(i),A} \quad (20)$$

$$G : \{1, \dots, N\} \mapsto \{1, \dots, K\} \quad (21)$$

where G is the map between stocks and clusters. The matrix Λ_{iA} is referred to as the loadings matrix. The Kronecker delta $\delta_{ab} = 1$ if $a = b$, and $\delta_{ab} = 0$ if $a \neq b$.

Mean-reversion can be done separately for each cluster as the clusters do not overlap. However, for further generalization, it is convenient to write the demeaned returns in a compact form, for all clusters at once. This brings in regression.

2.5 Regression

Let R_i be the stock returns. Consider a linear regression of R_i over Λ_{iA} (without intercept and with unit weights – see below). In R notation:⁸

$$R \sim -1 + \Lambda \quad (22)$$

where, in matrix notation, R is the N -vector R_i , and Λ is the $N \times K$ loadings matrix Λ_{iA} . Explicitly, we have

$$R_i = \sum_{A=1}^K \Lambda_{iA} f_A + \varepsilon_i \quad (23)$$

where f_A are the regression coefficients given by (in matrix notation)

$$f = Q^{-1} \Lambda^T R \quad (24)$$

$$Q \equiv \Lambda^T \Lambda \quad (25)$$

and ε_i are the regression residuals. In the case of binary Λ_{iA} we introduced in the previous subsection, these residuals are nothing but the returns R_i demeaned w.r.t. to the corresponding cluster:

$$\varepsilon = R - \Lambda Q^{-1} \Lambda^T R \quad (26)$$

$$Q_{AB} = N_A \delta_{AB} \quad (27)$$

$$\bar{R}_A \equiv \frac{1}{N_A} \sum_{j \in J_A} R_j \quad (28)$$

$$\varepsilon_i = R_i - \bar{R}_{G(i)} = \tilde{R}_i \quad (29)$$

⁸ The R Package for Statistical Computing. Also, “ \sim ” in (22) is R notation for a linear model.

where \bar{R}_A is the mean return for the cluster labeled by A , and \tilde{R}_i is the demeaned return obtained by subtracting from R_i the mean return for the cluster labeled by $A = G(i)$ to which the stock labeled by i belongs: $G(i) = A : i \in J_A \subset \{1, \dots, N\}$.

So, the demeaned returns \tilde{R}_i are given by the residuals of a regression (without intercept and with unit weights) of the returns R_i over the loadings matrix Λ_{iA} . This result allows to further generalize the above construction. But first some additional observations are in order.

Note that

$$\sum_{i=1}^N \tilde{R}_i \Lambda_{iA} = 0, \quad A = 1, \dots, K \quad (30)$$

I.e., the demeaned returns are cluster neutral, or, if the clusters are referred to as industries, they are industry neutral. In this case this is simply the statement that for each cluster the sum of the demeaned returns over all stocks in such cluster vanishes, which follows from the fact that the returns are demeaned w.r.t. each cluster. However, in a more general case (see below) this is a more nontrivial condition.

Also, note that we automatically have

$$\sum_{i=1}^N \tilde{R}_i \nu_i = 0 \quad (31)$$

where $\nu_i \equiv 1, i = 1, \dots, N$, *i.e.*, the N -vector ν is the unit vector. In the regression language, ν is referred to as the intercept. Above we did not have to add the intercept to the loadings matrix because it is already subsumed in it:

$$\sum_{A=1}^K \Lambda_{iA} = \nu_i \quad (32)$$

However, in the general case, to have (31), we would need to add the intercept as a column in the loadings matrix (see below). Recall that (31) is the same as dollar neutrality in the strategy (16), but generally dollar neutrality does not require (31).

2.6 Non-binary Generalization

The conditions (30) satisfied by the demeaned returns in the binary loadings matrix case simply mean that these returns are cluster neutral, *i.e.*, orthogonal to the corresponding N -vectors $v^{(A)}$, where $v_i^{(A)} \equiv \Lambda_{iA}$. That is, in matrix notation

$$\tilde{R}^T v^{(A)} = 0 \quad (33)$$

This orthogonality can be defined for any loadings matrix, not just a binary one.

This leads us to a generalization where the loadings matrix, call it Ω_{iA} , may have some binary columns, but generally it need not. The binary columns, if any,

are interpreted as industry (cluster) based risk factors; the non-binary columns are interpreted as some non-industry based risk factors; and the orthogonality condition

$$\sum_{i=1}^N \tilde{R}_i \Omega_{iA}, \quad A = 1, \dots, K \quad (34)$$

is simply the requirement that the twiddled returns \tilde{R}_i – which we will no longer refer to as “demeaned returns” (for the loadings matrix is no longer necessarily binary), but instead we will refer to them as “regressed returns” – are the residuals of the regression (without intercept and with unit weights) of R_i over Ω_{iA} :

$$\tilde{R} \equiv R - \Omega Q^{-1} \Omega^T R \quad (35)$$

$$Q \equiv \Omega^T \Omega \quad (36)$$

Note that we no longer necessarily have the property (31). If this property is desired, it can be achieved by including the intercept in the regression. *I.e.*, in the R notation the regression (now with intercept, but still with unit weights) is

$$R \sim \Omega \quad (37)$$

In terms of (35) this simply amounts to adding a unit column to Ω , so now we have $\Omega_{iA_1} \equiv \nu_i = 1, i = 1, \dots, N$ for some column labeled by A_1 .

2.7 Weighted Regression

In Subsection 2.3 we discussed a simple strategy (16), where the desired dollar holdings D_i are proportional to \tilde{R}_i . One potential “shortcoming” in this strategy is that on average its positions will be dominated by volatile stocks. One idea for reducing this exposure to volatility is to divide \tilde{R}_i by σ_i or σ_i^2 (or some other power of σ_i), where σ_i is, *e.g.*, the historical volatility of \tilde{R}_i or, more simply,⁹ of R_i , *i.e.*, variances $\sigma_i^2 \equiv C_{ii}$ are the diagonal elements of the sample covariance matrix

$$C_{ij} \equiv \langle R_i, R_j \rangle \quad (38)$$

where the covariances $\langle \cdot, \cdot \rangle$ are computed over the corresponding time series of R_i .

Here a few remarks are in order. First, if we take, say, $\hat{R}_i \equiv \tilde{R}_i / \sigma_i^2$, even if $\tilde{R}^T \nu = 0$, generally we do not have $\hat{R}^T \nu = 0$, so using \hat{R}_i instead of \tilde{R}_i in (16) would generally spoil the dollar neutrality property (*i.e.*, that $\sum_{i=1}^N D_i = 0$). We need to deal with this somehow. Second, should we take $\hat{R}_i \equiv \tilde{R}_i / \sigma_i$ or $\hat{R}_i \equiv \tilde{R}_i / \sigma_i^2$ or something else? The preferred answer to the last question is that we need to take $\hat{R}_i \equiv \tilde{R}_i / \sigma_i^2$ – or, more precisely, its variant we will come to in a moment – and the reason for this will become clear when we discuss optimization. This may appear

⁹ We will discuss these subtleties below, when we discuss factor models and optimization.

a bit odd at first as the return with the risk scaled out of it should be \tilde{R}_i/σ_i , not \tilde{R}_i/σ_i^2 . However, the extra suppression by another factor of σ_i is what maximizes the Sharpe ratio, which we will discuss in more detail below. For now, we will take this for granted and suppress the return \tilde{R}_i by σ_i^2 . All we have to figure out is how to make sure that we do not spoil dollar neutrality in the process.

One answer is given by weighted regression, where R is regressed over Ω with weights z_i . We have

$$\varepsilon \equiv R - \Omega Q^{-1} \Omega^T Z R \quad (39)$$

$$Z \equiv \text{diag}(z_i) \quad (40)$$

$$Q \equiv \Omega^T Z \Omega \quad (41)$$

$$\tilde{R} \equiv Z \varepsilon \quad (42)$$

Here ε_i are the residuals of the weighted regression. Also, note that

$$\sum_{i=1}^N \tilde{R}_i \Omega_{iA} = 0, \quad A = 1, \dots, K \quad (43)$$

If the intercept is included in Ω_{iA} , then we automatically have $\sum_{i=1}^N \tilde{R}_i = 0$. Also, if we take $z_i = 1/\sigma_i^2$, then \tilde{R}_i will be suppressed by σ_i^2 compared with the case of the regression with unit weights. So, now our simple strategy (16) is not only dollar neutral but has risk management built into it. The resulting holdings are neutral w.r.t. the risk factors described by the loadings matrix Ω_{iA} , and furthermore are no longer dominated by volatile stock holdings. This is now a real mean-reversion strategy. We will take it a step further in the next section.

2.8 Remarks

As mentioned above, (16) is only one of a myriad ways of specifying D_i given the regressed returns \tilde{R}_i . If the regression includes the intercept, \tilde{R}_i have 0 cross-sectional mean, so the strategy defined by (16) is automatically dollar neutral. Also, if the regression is weighted as in Subsection 2.7, contributions from high volatility stocks are weighted down thereby providing risk management.¹⁰ Here, for illustrative purposes only (and not as an exhaustive survey), we discuss some other mean-reversion strategies, *i.e.*, other ways of specifying D_i .

One simple example is to have “equally weighted” D_i

$$D_i = -\gamma \text{sign}(\tilde{R}_i) \quad (44)$$

where $\gamma > 0$, *i.e.*, we buy stocks with negative regressed returns and sell stocks with positive regressed returns, all with the same absolute dollar amount equal $\gamma = I/N$

¹⁰ As we discuss in the next section, this case is a certain limit of optimization.

(this equality follows from (13)). This strategy has some evident “shortcomings”. First, it is not necessarily dollar neutral:

$$\sum_{i=1}^N D_i = \frac{N_- - N_+}{N} I \quad (45)$$

where N_+ is the number of stocks with positive regressed returns and N_- is the number of stocks with negative regressed returns, and generally $N_+ \neq N_-$. If N is large, then *assuming* a normal distribution for $N_- - N_+$ with mean 0 and standard deviation of order \sqrt{N} , the mishedge (45) is of order I/\sqrt{N} . For $N \sim 2,500$, this is of order 2%, which may be unacceptably large. To achieve dollar neutrality, we can modify the values of some D_i , *e.g.*, by setting some of them to zero. One then needs to decide which values to set to zero. This brings us to the second “shortcoming” in this strategy: $\text{sign}(x)$ is discontinuous across $x = 0$, so for small \tilde{R}_i the sign of D_i can flip even with small fluctuations.¹¹ This instability can result in unnecessary portfolio turnover (overtrading) and additional trading costs, and generally diminish the performance of the strategy. One way to “smooth” this out is to approximate $\text{sign}(x)$ via, *e.g.*, $\tanh(x/\kappa)$:

$$D_i = -\gamma \tanh(\tilde{R}_i/\kappa) \quad (46)$$

where κ is the *cross-sectional* standard deviation of \tilde{R}_i . Then, for $|\tilde{R}_i| \ll \kappa$ this approximately reduces to (16), whereas for $|\tilde{R}_i| \gtrsim \kappa$ the dollar holdings are “squashed”. If the regression has unit weights, then on average $|\tilde{R}_i|$ are larger for more volatile stocks compared with less volatile stocks, and using (46) amounts to suppressing the contributions from less volatile stocks while “equally” weighting the contributions from more volatile stocks. As mentioned above, this may not be desirable from the risk management viewpoint. If the regression is weighted with $z_i = 1/\sigma_i^2$, then on average $|\tilde{R}_i|$ are suppressed for more volatile stocks compared with less volatile stocks, so using (46) amounts to suppressing the contributions from more volatile stocks, while “equally” weighting the contributions from less volatile stocks. In this case we can achieve dollar neutrality by setting to zero and/or appropriately scaling down the absolute values of D_i for more volatile stocks. Let us mention that (46) generally is farther away than (16) (assuming \tilde{R}_i are based on a weighted regression with $z_i = 1/\sigma_i^2$) from the optimized solution we discuss in the next section.

If one contemplates (44) (or (46) as its “smoothed out” version), one may also explore the opposite direction and consider, *e.g.*,

$$D_i = -\gamma \tilde{R}_i |\tilde{R}_i| \quad (47)$$

Here \tilde{R}_i are based on a weighted regression – otherwise the portfolio would be too volatile. In fact, more generally one can consider strategies with

$$D_i = -\gamma \tilde{R}_i f(\tilde{R}_i) \quad (48)$$

¹¹ Be it due to changes from one day to another, or due to computational uncertainties, *etc.*

where $f(x)$ is some function. Such “nonlinear alphas” are commonly used in quant trading. Note that, as for (44) and (46), (47) and more generally (48) require additional “gymnastics” to achieve dollar neutrality. There is no “magic prescription” for picking $f(x)$ in (48). In practice at any given time one picks alphas that backtest well, and alphas are ephemeral by nature – alphas that work now may not work 6 months from now. This is an ever-changing empirical game, not a theoretical one.

This brings us to yet another commonly used way of specifying D_i : ranking. Instead of using continuous functions such as, *e.g.*, (16) or (48), one can, *e.g.*, rank stocks cross-sectionally by $|\tilde{R}_i|$. Let this integer rank be r_i . Then we can take, *e.g.*:

$$D_i = -\gamma \operatorname{sign}(\tilde{R}_i) r_i \quad (49)$$

Alternatively, we can set D_i to 0 for the stocks with $r_i < r_*$. Various comments we made above relating to risk management and dollar neutrality also apply to alphas based on ranking. Furthermore, one can consider nonlinear functions of r_i .

In this regard, let us also mention that above we treat dollar neutrality symmetrically between long and short holdings. There are other possibilities here too. *E.g.*, we can go long cash (*i.e.*, stocks – more trader lingo) with D_i specified via, say, (16), and short the same dollar amount of futures for some diversified index, *e.g.* S&P500 – this would be a so-called S&P outperformance portfolio. In this case we have lower bounds $D_i \geq 0$. Similarly, instead of shorting futures, we could short a tracking portfolio for the index, *e.g.*, a minimum variance portfolio, whose weights are independent of the stock expected returns.¹² Instead of limiting the short position to a tracking portfolio for an index, one can consider a minimum variance portfolio for some (diversified) proprietary trading universe. As mentioned above, there are many ways of doing mean-reversion. Here we focus on the sequence “mean-reversion via demeaning \rightarrow regression \rightarrow weighted regression \rightarrow (constrained) optimization \rightarrow factor models”, which brings us to our next topic – optimization.

3 Optimization

3.1 Maximizing Sharpe Ratio

Let C_{ij} be the sample covariance matrix of the N time series of stock returns $R_i(t_s)$, $s = 0, 1, \dots, M$, where t_0 is the most recent time. Below R_i refers to $R_i(t_0)$. Let Ψ_{ij} be the corresponding correlation matrix, *i.e.*,

$$C_{ij} = \sigma_i \sigma_j \Psi_{ij} \quad (50)$$

where $\Psi_{ii} = 1$. For the sake of definiteness, let us assume that R_i are daily returns, albeit this is not a critical assumption.

¹² In this case, the actual portfolio consists of net long or short positions for individual stocks arising from long positions D_i and short positions from the minimum variance portfolio.

As above, let D_i be the dollar holdings in our portfolio. The portfolio P&L, volatility and Sharpe ratio are given by

$$P = \sum_{i=1}^N R_i D_i \quad (51)$$

$$V = \sqrt{\sum_{i,j=1}^N C_{ij} D_i D_j} \quad (52)$$

$$S = \frac{P}{V} \quad (53)$$

Instead of dollar holdings D_i , it is more convenient to work with dimensionless holding weights (not to be confused with the regression weights z_i)

$$w_i \equiv \frac{D_i}{I} \quad (54)$$

where I is the investment level. The holding weights satisfy the condition

$$\sum_{i=1}^N |w_i| = 1 \quad (55)$$

They are positive for long holdings and negative for short holdings.

In terms of the holding weights, the P&L and volatility are given by

$$P = I \tilde{P} \equiv I \sum_{i=1}^N R_i w_i \quad (56)$$

$$V = I \tilde{V} \equiv I \sqrt{\sum_{i,j=1}^N C_{ij} w_i w_j} \quad (57)$$

To determine the weights, often one requires that the Sharpe ratio be maximized:

$$S \rightarrow \max \quad (58)$$

Assuming (for now) that there are no additional conditions on w_i (*e.g.*, upper or lower bounds), the solution to (58) in the absence of costs is given by

$$w_i = \gamma \sum_{j=1}^N C_{ij}^{-1} R_j \quad (59)$$

where C^{-1} is the inverse of C , and the normalization coefficient γ is determined from (55). Invertibility of C should not be taken for granted and we will discuss this issue a bit later. However, for now, let us assume that C is invertible.

One immediate consequence of (59) is that these holding weights generically do not correspond to a dollar neutral portfolio. *E.g.*, if C_{ij} is diagonal and all $R_i > 0$, then all $w_i > 0$. More generally, there is no reason why $\sum_{i=1}^N w_i$ should vanish. So, if we wish to have a dollar neutral portfolio, we need to maximize the Sharpe ratio subject to the dollar neutrality constraint.

3.2 Linear Constraints; Dollar Neutrality

Dollar neutrality can be achieved as follows. First, note that the Sharpe ratio is invariant under the simultaneous rescalings of all holding weights $w_i \rightarrow \zeta w_i$, where $\zeta > 0$. Because of this scale invariance, the Sharpe ratio maximization problem can be recast in terms of minimizing a quadratic objective function:

$$g(w, \lambda) \equiv \frac{\lambda}{2} \sum_{i,j=1}^N C_{ij} w_i w_j - \sum_{i=1}^N R_i w_i \quad (60)$$

$$g(w, \lambda) \rightarrow \min \quad (61)$$

where $\lambda > 0$ is a parameter, and minimization is w.r.t. w_i . The solution is given by

$$w_i = \frac{1}{\lambda} \sum_{j=1}^N C_{ij}^{-1} R_j \quad (62)$$

and λ is fixed via (55). The objective function approach is convenient if we wish to impose constraints on w_i , *e.g.*, the dollar neutrality constraint. We introduce an $N \times m$ matrix Y_{ia} and m Lagrange multipliers μ_a , $a = 1, \dots, m$:

$$g(w, \mu, \lambda) \equiv \frac{\lambda}{2} \sum_{i,j=1}^N C_{ij} w_i w_j - \sum_{i=1}^N R_i w_i - \sum_{a=1}^m \sum_{i=1}^N w_i Y_{ia} \mu_a \quad (63)$$

$$g(w, \mu, \lambda) \rightarrow \min \quad (64)$$

Minimization w.r.t. w_i and μ_a now gives the following equations:

$$\lambda \sum_{j=1}^N C_{ij} w_j = R_i + \sum_{a=1}^m Y_{ia} \mu_a \quad (65)$$

$$\sum_{i=1}^N w_i Y_{ia} = 0 \quad (66)$$

So, we have m homogeneous linear constraints (66). If $Y_{ia_1} \equiv \nu_i = 1$, $i = 1, \dots, N$ for some $a_1 \in \{1, \dots, m\}$, then we have dollar neutrality. Note that m can be 1.

The solution to (65) and (66) is given by (in matrix notation):

$$w = \frac{1}{\lambda} [C^{-1} - C^{-1} Y (Y^T C^{-1} Y)^{-1} Y^T C^{-1}] R \quad (67)$$

$$\mu = -(Y^T C^{-1} Y)^{-1} Y^T C^{-1} R \quad (68)$$

As before, λ is fixed via (55). The solution (67) and (68) can be rewritten as follows:

$$\omega = \frac{1}{\lambda} \Gamma^{-1} \rho \quad (69)$$

$$\omega^T \equiv (w^T, -\lambda^{-1} \mu^T) \quad (70)$$

$$\rho^T \equiv (R^T, O^T) \quad (71)$$

$$\Gamma \equiv \begin{pmatrix} C & Y \\ Y^T & \mathbb{O} \end{pmatrix} \quad (72)$$

I.e., ω and ρ are $(N+m)$ -vectors, and Γ is an $(N+m) \times (N+m)$ matrix; O is a nil m -vector, and \mathbb{O} is a nil $m \times m$ matrix. Thus, linear constraints can be dealt with by simply enlarging the covariance matrix as above.¹³

3.3 Regression as Constrained Diagonal Optimization

Let us now consider the case where the covariance matrix is diagonal: $C_{ij} = \sigma_i^2 \delta_{ij}$. Then (67) reads

$$w = \frac{1}{\lambda} [Z - Z Y (Y^T Z Y)^{-1} Y^T Z] R = \frac{1}{\lambda} Z \varepsilon = \frac{1}{\lambda} \tilde{R} \quad (73)$$

Here $Z \equiv \text{diag}(1/\sigma_i^2)$, ε_i are the residuals of the weighted regression with weights $z_i = 1/\sigma_i^2$ of R_i over the $N \times m$ matrix Y_{ia} (without intercept – unless the intercept is already included in Y_{ia} , that is). This is the same weighted regression we discussed in Subsection 2.7. So, diagonal (meaning, with diagonal covariance matrix) constrained optimization is the same as the weighted regression with the loadings matrix Ω identified with the constraint matrix Y , and the regression weights z_i (not to be confused with the holding weights w_i) identified with inverse variances of the returns R . Furthermore, the holding weights w_i are given by the regressed returns \tilde{R}_i up to a normalization factor fixed via (55). If the constraint matrix contains the intercept (the unit vector), then the holding weights correspond to a dollar neutral portfolio.

3.4 Regression as Limit of Optimization

Weighted regression (39), (40), (41) and (42) has a structure such that it is actually related to factor models. Consider an auxiliary matrix

$$\Theta \equiv \Xi + \zeta \Omega \Omega^T \quad (74)$$

$$\Xi \equiv Z^{-1} \quad (75)$$

where ζ is a parameter. The inverse reads:

$$\Theta^{-1} = Z - \zeta Z \Omega \tilde{Q}^{-1} \Omega^T Z \quad (76)$$

$$\tilde{Q}_{AB} \equiv \delta_{AB} + \zeta \sum_{i=1}^N z_i \Omega_{iA} \Omega_{iB} \quad (77)$$

¹³ Above we considered homogeneous constraints (66). Technically, the same trick can be applied to inhomogeneous constraints of the form $\sum_{i=1}^N w_i Y_{ia} + y_a = 0$. Everything goes through as above, except that now we have $\rho_a = -\lambda y_a$. However, while this will give the correct solution to the minimization of the objective function, this is no longer necessarily the same as maximizing the Sharpe ratio with constraints: the latter explicitly break the invariance under the rescalings $w_i \rightarrow \zeta w_i$ (unless $y_a \equiv 0$), which is what allowed us to rewrite the Sharpe ratio maximization problem in terms of the objective function minimization problem, whereby λ is fixed via (55). In the presence of inhomogeneous constraints this is no longer the case and some additional care is needed – see Section 5. We will not need inhomogeneous constraints here, however.

In the $\zeta \rightarrow \infty$ limit, which (with some care) can be thought of as a $1/z_i \rightarrow 0$ limit, we have

$$\Theta^{-1} = Z - Z \Omega Q^{-1} \Omega^T Z \quad (78)$$

$$Q \equiv \Omega^T Z \Omega \quad (79)$$

$$\tilde{R} = \Theta^{-1} R \quad (80)$$

where \tilde{R} is the vector of regressed returns in (42). So, regression is indeed a limit of optimization where the covariance matrix is given by Θ . This is the factor model form – with a subtlety, that is (see below).

3.5 Factor Models

In a multi-factor risk model, instead of N stock returns R_i , one deals with $K \ll N$ risk factors and the covariance matrix C_{ij} is replaced by Θ_{ij} given by

$$\Theta \equiv \Xi + \tilde{\Omega} \Phi \tilde{\Omega}^T \quad (81)$$

$$\Xi_{ij} \equiv \xi_i^2 \delta_{ij} \quad (82)$$

where ξ_i is the specific (a.k.a. idiosyncratic) risk for each stock; $\tilde{\Omega}_{iA}$ is an $N \times K$ factor loadings matrix; and Φ_{AB} is the factor covariance matrix, $A, B = 1, \dots, K$. *I.e.*, the random processes Υ_i corresponding to N stocks are modeled via N random processes χ_i (corresponding to specific risk) together with K random processes f_A (corresponding to factor risk):

$$\Upsilon_i = \chi_i + \sum_{A=1}^K \tilde{\Omega}_{iA} f_A \quad (83)$$

$$\langle \chi_i, \chi_j \rangle = \Xi_{ij} \quad (84)$$

$$\langle \chi_i, f_A \rangle = 0 \quad (85)$$

$$\langle f_A, f_B \rangle = \Phi_{AB} \quad (86)$$

$$\langle \Upsilon_i, \Upsilon_j \rangle = \Theta_{ij} \quad (87)$$

Instead of an $N \times N$ covariance matrix C_{ij} we now have a $K \times K$ factor covariance matrix Φ_{AB} . We have

$$\Theta = \Xi + \Omega \Omega^T \quad (88)$$

$$\Omega \equiv \tilde{\Omega} \tilde{\Phi} \quad (89)$$

$$\tilde{\Phi} \tilde{\Phi}^T = \Phi \quad (90)$$

where $\tilde{\Phi}_{AB}$ is the Cholesky decomposition of Φ_{AB} , which is assumed to be positive-definite. Note that, in the notations of the previous subsection, we have chosen the normalization such that $\zeta = 1$.

In the factor model approach, one replaces the sample covariance matrix C_{ij} (which is computed based on the time series of the returns R_i) by Θ_{ij} . The main reason for doing so is that the off-diagonal elements of C_{ij} typically are not expected to be too stable out-of-sample. In this regard, a constructed factor model covariance matrix Θ_{ij} is expected to be much more stable. This is because the number of factors, for which the factor covariance matrix Φ_{AB} needs to be computed, is $K \ll N$. Furthermore, if $M < N$ (recall that $M + 1$ is the number of observations in each time series), then C_{ij} is singular – it has only $M < N$ nonzero eigenvalues in this case. Note that, assuming all specific risks $\xi_i > 0$ and the factor covariance matrix Φ_{AB} is positive-definite, then Θ_{ij} is automatically positive-definite (and invertible).

3.6 Optimization with Factor Model

So, suppose we have a factor model covariance matrix Θ_{ij} . If we maximize the Sharpe ratio using this factor model covariance matrix, the resulting holding weights are given by (λ is fixed via (55))

$$w_i = \frac{1}{\lambda \xi_i^2} \left(R_i - \sum_{j=1}^N \frac{R_j}{\xi_j^2} \sum_{A,B=1}^K \Omega_{iA} \Omega_{jB} \tilde{Q}_{AB}^{-1} \right) \quad (91)$$

$$\tilde{Q}_{AB} \equiv \delta_{AB} + \sum_{i=1}^N \frac{1}{\xi_i^2} \Omega_{iA} \Omega_{iB} \quad (92)$$

where \tilde{Q}_{AB}^{-1} is the inverse of \tilde{Q}_{AB} . As in the general case, these holding weights are not dollar neutral.

3.6.1 Linear Constraints

As in the general case, in the factor model context too we can incorporate multiple (homogeneous) linear constraints (66). Let $\hat{\Omega}_{i\alpha}$, $\alpha \in H \equiv \{a\} \cup \{A\}$ (*i.e.*, the index α has m values corresponding to the index a and K values corresponding to the index A) be the following $N \times (K + m)$ matrix:

$$\hat{\Omega}_{ia} \equiv Y_{ia} \quad (93)$$

$$\hat{\Omega}_{iA} \equiv \Omega_{iA} \quad (94)$$

The corresponding holding weights then are given by:

$$w_i = \frac{1}{\lambda \xi_i^2} \left(R_i - \sum_{j=1}^N \frac{R_j}{\xi_j^2} \sum_{\alpha, \beta \in H} \hat{\Omega}_{i\alpha} \hat{\Omega}_{j\beta} \hat{Q}_{\alpha\beta}^{-1} \right) \quad (95)$$

$$\hat{Q}_{\alpha\beta} \equiv \varphi_{\alpha\beta} + \sum_{i=1}^N \frac{1}{\xi_i^2} \hat{\Omega}_{i\alpha} \hat{\Omega}_{i\beta} \quad (96)$$

where $\widehat{Q}_{\alpha\beta}^{-1}$ is the inverse of $\widehat{Q}_{\alpha\beta}$, and $\varphi_{AB} = \delta_{AB}$, $\varphi_{ab} = \varphi_{aA} = \varphi_{Aa} = 0$. We have

$$\sum_{i=1}^N w_i Y_{ia} = \sum_{i=1}^N w_i \widehat{\Omega}_{ia} = \sum_{\alpha, \beta \in H} \sum_{j=1}^N \frac{R_j}{\xi_j^2} \widehat{\Omega}_{j\beta} \varphi_{a\alpha} \widehat{Q}_{\alpha\beta}^{-1} = 0 \quad (97)$$

So, the holding weights w_i satisfy the constraints (66).

3.6.2 Optimization with Constraints

The constraints (66) typically are related to risk management. Apart from dollar neutrality (*i.e.*, roughly, the market neutrality constraint), other constraints typically are the requirements of neutrality w.r.t. other risk factors, *e.g.*, industry neutrality, neutrality w.r.t. style risk factors (*e.g.*, size, liquidity, volatility, momentum, *etc.*) or other non-industry risk factors (*e.g.*, principal component based risk factors or betas). In practice, one often uses the same risk factors in Y_{ia} as those in the factor loadings matrix Ω_{iA} in the factor model.¹⁴ If that is the case, then there is certain redundancy in the matrix $\widehat{\Omega}_{i\alpha}$, which we turn to next.

Since we can always rotate the constraints (66) by an arbitrary non-singular $m \times m$ matrix, we can separate these constraints into two sets, $\{a\} = \{a'\} \cup \{a''\} \equiv J' \cup J''$, such that $Y_{ia''}$ are “orthogonal” to Ω_{iA} and no further rotation can make $Y_{ia'}$ “orthogonal” to Ω_{iA} :

$$\sum_{i=1}^N \frac{1}{\xi_i^2} \Omega_{iA} Y_{ia''} = 0, \quad A = 1, \dots, K, \quad a'' \in J'' \quad (98)$$

Let us assume J'' is not empty – if it is empty, we can still proceed as below, except that $\epsilon_i'' = R_i$ in this case (see below).

Let $H' \equiv \{A\} \cup J' = H \setminus J''$ (recall that $H = \{A\} \cup \{a\}$). Then we have

$$\begin{aligned} w_i &= \frac{1}{\lambda \xi_i^2} \left(R_i - \sum_{j=1}^N \frac{R_j}{\xi_j^2} \sum_{\alpha, \beta \in H} \widehat{\Omega}_{i\alpha} \widehat{\Omega}_{j\beta} \widehat{Q}_{\alpha\beta}^{-1} \right) = \\ &= \frac{1}{\lambda \xi_i^2} \left(\epsilon_i'' - \sum_{j=1}^N \frac{\epsilon_j''}{\xi_j^2} \sum_{\alpha', \beta' \in H'} \widehat{\Omega}_{i\alpha'} \widehat{\Omega}_{j\beta'} \widehat{Q}_{\alpha'\beta'}^{-1} \right) \end{aligned} \quad (99)$$

where

$$\epsilon_i'' \equiv R_i - \sum_{j=1}^N \frac{R_j}{\xi_j^2} \sum_{a'', b'' \in J''} Y_{ia''} Y_{jb''} Q_{a''b''}^{-1} \quad (100)$$

$$Q_{a''b''} \equiv \sum_{i=1}^N \frac{1}{\xi_i^2} Y_{ia''} Y_{ib''} \quad (101)$$

¹⁴ More precisely, usually one would use the unrotated factor loadings $\widetilde{\Omega}_{iA}$ – recall that $\Omega = \widetilde{\Omega} \widetilde{\Phi}$, where $\widetilde{\Phi}$ is the Cholesky decomposition of the factor covariance matrix Φ . However, a rotation $Y \rightarrow YU$ by an arbitrary nonsingular $m \times m$ matrix U_{ab} does not change the constraints (66).

and $Q_{a''b''}^{-1}$ is the inverse of the $|J''| \times |J''|$ matrix $Q_{a''b''}$, $a'', b'' \in J''$.

So, ε_i'' are nothing but the regression residuals of R_i regressed over $Y_{ia''}$ with regression weights $z_i'' \equiv 1/\xi_i^2$. Put differently, our original constrained optimization has reduced to constrained optimization with a subset of the original constraints

$$\sum_{i=1}^N w_i Y_{ia'} = 0, \quad a' \in J' \quad (102)$$

but instead of optimizing the returns R_i , we are now optimizing the regression residuals ε_i'' . This is because the original matrix $\hat{Q}_{\alpha\beta}$ is block-diagonal:

$$\hat{Q}_{\alpha'a''} = 0, \quad \alpha' \in H', \quad a'' \in J'' \quad (103)$$

$$\hat{Q}_{a''b''} = Q_{a''b''}, \quad a'', b'' \in J'' \quad (104)$$

In fact, we can break this down further.

Let us assume that the $|J'|$ columns in the remaining loadings $Y_{ia'}$, $a' \in J'$ are a subset of the columns in the factor loadings Ω_{iA} . So we have $\{A\} = \{A'\} \cup J' \equiv F' \cup J'$, and the $K' \equiv |F'| = K - |J'|$ values of the index A' run over the columns in Ω_{iA} which differ from those in $Y_{ia'}$. Further, to avoid notational confusion, we will denote $\Omega_{iA}|_{A=A'} \equiv \Omega'_{iA'}$, $A' \in F'$. It is then not difficult to show that

$$\hat{Q}_{\alpha'\beta'}^{-1} = \begin{pmatrix} D^{-1} & \mathbb{O} & -D^{-1}E\Delta^{-1} \\ \mathbb{O} & \mathbb{I} & -\mathbb{I} \\ -\Delta^{-1}E^TD^{-1} & -\mathbb{I} & \mathbb{I} + \Delta^{-1} + \Delta^{-1}E^TD^{-1}E\Delta^{-1} \end{pmatrix} \quad (105)$$

Here \mathbb{I} is the $|J'| \times |J'|$ identity matrix, while \mathbb{O} is the $|F'| \times |J'|$ nil matrix, D is an $|F'| \times |F'|$ matrix, E is an $|F'| \times |J'|$ matrix and Δ is a $|J'| \times |J'|$ matrix:

$$D \equiv \tilde{Q}' - E\Delta^{-1}E^T \quad (106)$$

$$\tilde{Q}'_{A'B'} \equiv \delta_{A'B'} + \sum_{i=1}^N \frac{1}{\xi_i^2} \Omega'_{iA'} \Omega'_{iB'}, \quad A', B' \in F' \quad (107)$$

$$E_{A'b'} \equiv \sum_{i=1}^N \frac{1}{\xi_i^2} \Omega'_{iA'} Y_{ib'}, \quad A' \in F', \quad b' \in J' \quad (108)$$

$$\Delta_{a'b'} \equiv \sum_{i=1}^N \frac{1}{\xi_i^2} Y_{ia'} Y_{ib'}, \quad a', b' \in J' \quad (109)$$

We therefore have (in matrix notation – here Y refers to the $Y_{ia'}$ matrix)

$$\begin{aligned} w &= \frac{1}{\lambda} \{ \Xi^{-1} - \Xi^{-1} [\Omega' D^{-1} \Omega' - \Omega' D^{-1} E \Delta^{-1} Y^T - Y \Delta^{-1} E^T D^{-1} \Omega' + \\ &+ Y (\Delta^{-1} + \Delta^{-1} E^T D^{-1} E \Delta^{-1}) Y^T] \Xi^{-1} \} \varepsilon'' \end{aligned} \quad (110)$$

Furthermore

$$Y^T w = 0 \quad (111)$$

In fact, w_i given by (110) correspond to optimizing the residuals ε_i'' using a reduced factor model with the same specific risk but the factor loadings given by $\Omega'_{iA'}$

$$\Theta'_{ij} \equiv \xi_i^2 \delta_{ij} + \sum_{A'=1}^{K'} \Omega'_{iA'} \Omega'_{jA'} \quad (112)$$

subject to the constraints

$$\sum_{i=1}^N w_i Y_{ia'} = 0, \quad a' \in J' \quad (113)$$

The solution to this optimization problem is given by

$$w_i = \frac{1}{\lambda \xi_i^2} \left(\varepsilon_i'' - \sum_{j=1}^N \frac{\varepsilon_j''}{\xi_j^2} \sum_{\alpha^*, \beta^* \in F^*} \hat{\Omega}_{i\alpha^*} \hat{\Omega}_{j\beta^*} \hat{Q}_{\alpha^*\beta^*}^{-1} \right) \quad (114)$$

where $F^* \equiv F' \cup J'$ (so F^* as a set is the same as $\{A\}$, but we use a different notation for it to avoid confusion), and we have $\hat{\Omega}_{i\alpha^*} \Big|_{\alpha^*=A'} \equiv \Omega'_{iA'}$, $\hat{\Omega}_{i\alpha^*} \Big|_{\alpha^*=a'} \equiv Y_{ia'}$, and

$$\hat{Q}_{\alpha^*\beta^*}^{-1} = \begin{pmatrix} D^{-1} & -D^{-1}E\Delta^{-1} \\ -\Delta^{-1}E^T D^{-1} & \Delta^{-1} + \Delta^{-1}E^T D^{-1}E\Delta^{-1} \end{pmatrix} \quad (115)$$

It then follows that w_i in (110) and (114) are identical.

To summarize, optimization is done with the columns in the factor loadings matrix Ω_{iA} corresponding to the columns in Y_{ia} omitted.

3.7 Pitfalls

So, what happens if we run constrained optimization with the factor loadings matrix as Y_{ia} ? *I.e.*, $m = K$, the index a takes the same values as the index A , and $Y_{ia}|_{a=A} = \Omega_{iA}$. (If we wish to have dollar neutrality, we simply assume that Ω_{iA} contains the intercept.) In this case, using the results of the previous subsection:

$$w_i = \frac{1}{\lambda \xi_i^2} \left(R_i - \sum_{j=1}^N \frac{R_j}{\xi_j^2} \sum_{A,B=1}^K \Omega_{iA} \Omega_{iB} Q_{AB}^{-1} \right) \quad (116)$$

$$Q_{AB} \equiv \sum_{i=1}^N \frac{1}{\xi_i^2} \Omega_{iA} \Omega_{iB} \quad (117)$$

so w_i are the same as in the weighted regression with regression weights $z_i = 1/\xi_i^2$.

3.7.1 Specific Risk or Total Risk?

In the optimization context, the regression weights in (116) naturally come out to be $z_i = 1/\xi_i^2$, the inverse of specific volatility squared, not total volatility (*i.e.*, $z_i \neq 1/\sigma_i^2$). This addresses the subtlety mentioned at the end of Subsection 3.4. However, *a priori* there is nothing wrong with using $z_i = 1/\sigma_i^2$ in the weighted regression outside of the factor model context. Specific risk is not known unless a factor model is available or carefully constructed. In that case, total risk is what is available for using as the regression weights, and typically can be so used.

3.7.2 Optimization of Regression Residuals

Instead of imposing constraints in optimization, one may be tempted to first regress returns R_i over (some) factor loadings Ω_{iA} to obtain regression residuals ε_i , and do the optimization based on these residuals (as opposed to the returns R_i). The rationale here is that the regressed returns $\tilde{R}_i = z_i \varepsilon_i$ (where z_i are the regression weights) are neutral w.r.t. the loadings used in the regression. However, unless this is done correctly, the resulting holding weights will not be neutral w.r.t. Ω_{iA} as the optimization in general undoes any such neutrality.

Thus, consider the following strategy:

$$\varepsilon_i \equiv R - Y (Y^T Z Y)^{-1} Y^T Z R \quad (118)$$

$$Z \equiv \text{diag}(z_i) \quad (119)$$

$$w_i \equiv \frac{1}{\lambda \xi_i^2} \left(\varepsilon_i - \sum_{j=1}^N \frac{\varepsilon_j}{\xi_j^2} \sum_{A,B=1}^K \Omega_{iA} \Omega_{jB} \tilde{Q}_{AB}^{-1} \right) \quad (120)$$

$$\tilde{Q}_{AB} \equiv \delta_{AB} + \sum_{i=1}^N \frac{1}{\xi_i^2} \Omega_{iA} \Omega_{iB} \quad (121)$$

Here we have purposefully kept the loadings Y_{ia} in the weighted regression (first equation above) distinct from the factor loadings Ω_{iA} in the optimization (third equation above). Note that the optimization is done on the regression residuals ε_i , not on the returns R_i .

If the purpose of the regression is that the holding weights be neutral w.r.t. Y_{ia} , then to ensure this, the matrix Y_{ia} must be the same as Ω_{iA} (modulo immaterial rotations – see footnote 14), *i.e.*, $m = K$, and the regression weights z_i cannot be arbitrary but must be taken as inverse specific variances: $z_i = 1/\xi_i^2$. Indeed, from (118) we then have

$$\sum_{i=1}^N \tilde{R}_i \Omega_{iA} = 0 \quad (122)$$

$$\tilde{R}_i \equiv z_i \varepsilon_i = \frac{1}{\xi_i^2} \varepsilon_i \quad (123)$$

so that

$$w_i = \frac{1}{\lambda} \tilde{R}_i \quad (124)$$

$$\sum_{i=1}^N w_i \Omega_{iA} = 0 \quad (125)$$

I.e., optimization of regression residuals (up to an overall proportionality constant $1/\lambda$) simply reduces to the regressed returns \tilde{R}_i , which are neutral w.r.t. Ω_{iA} .

4 “Intermezzo”

In Section 2 we started with simple pair trading and by the end of the last section it got substantially more involved. This trend is going to continue in the following sections, so this is a good place for an “intermezzo”. We will try to keep it light.

So, consider two stocks, A and B. Let their (sample) covariance matrix be

$$C = \begin{pmatrix} \sigma_A^2 & \rho \sigma_A \sigma_B \\ \rho \sigma_A \sigma_B & \sigma_B^2 \end{pmatrix} \quad (126)$$

Here σ_A and σ_B are the volatilities, and ρ is the correlation. Let our portfolio have D_A and D_B dollar holdings in A and B. Let R_A and R_B be the expected returns for A and B. Then the expected Sharpe ratio of this portfolio is

$$S = \frac{D_A R_A + D_B R_B}{\sqrt{(\sigma_A D_A)^2 + (\sigma_B D_B)^2 + 2 \rho (\sigma_A D_A)(\sigma_B D_B)}} \quad (127)$$

It is maximized by

$$D_A = \gamma \left(\frac{R_A}{\sigma_A^2} - \frac{\rho R_B}{\sigma_A \sigma_B} \right) \quad (128)$$

$$D_B = \gamma \left(\frac{R_B}{\sigma_B^2} - \frac{\rho R_A}{\sigma_A \sigma_B} \right) \quad (129)$$

Here $\gamma > 0$ is an arbitrary constant, which is a consequence of the invariance of S under simultaneous rescalings $D_A \rightarrow \zeta D_A$, $D_B \rightarrow \zeta D_B$ ($\zeta > 0$), and it is fixed via the requirement that $|D_A| + |D_B| = I$, where I is the investment level.

Now let us assume that the volatilities are the same: $\sigma_A = \sigma_B \equiv \sigma$. We have

$$D_A = \frac{\gamma}{\sigma^2} (R_A - \rho R_B) \quad (130)$$

$$D_B = \frac{\gamma}{\sigma^2} (R_B - \rho R_A) \quad (131)$$

As $\rho \rightarrow 1$, we have $D_A + D_B \rightarrow 0$, which is the dollar neutrality condition. So, the $S \rightarrow \max$ optimization, in the limit where volatilities are identical and the correlation goes to 1, produces a dollar neutral portfolio. How come? Two answers.

First, when $\sigma_A = \sigma_B$ and $\rho = 1$, the covariance matrix C is singular. The eigenvector V_i corresponding to the null eigenvalue is $V^T = (1, -1)$, and in this direction the portfolio volatility vanishes and the Sharpe ratio goes to infinity (see below). This is why $D_B = -D_A$ maximizes the Sharpe ratio.¹⁵

Second, we can tie this to Subsection 3.4. Consider a *one*-factor model for two stocks A and B : $\Theta = \Xi + \zeta \Omega \Omega^T$, where $\Xi = \text{diag}(\xi_A^2, \xi_B^2)$, and $\Omega^T = (1, 1)$. Then $\Theta = C$ in (126) with $\sigma_{A,B}^2 = \xi_{A,B}^2 + \zeta$, and $\rho = \zeta/(\sigma_A \sigma_B)$. In the $\zeta \rightarrow \infty$ limit (with ξ_A and ξ_B fixed) we have $\sigma_{A,B}^2 \rightarrow \zeta \equiv \sigma^2$, and $\rho \rightarrow 1$, exactly as above. On the other hand, as we saw in Subsection 3.4, in this limit optimization reduces to a regression over Ω , which is nothing but the intercept, hence dollar neutrality.

5 Optimization with Costs

5.1 Linear Costs

Above we ignored trading costs. Let us start by adding linear costs:¹⁶

$$P = \sum_{i=1}^N R_i D_i - \sum_{i=1}^N \tilde{L}_i |D_i - D_i^*| \quad (132)$$

where \tilde{L}_i for each stock includes, per each *dollar* traded, all fixed trading costs (SEC fees, exchange fees, broker-dealer fees, *etc.*) and linear slippage. The linear cost assumes no impact, *i.e.*, trading does not affect the stock prices. Also, D_i are the desired dollar holdings, and D_i^* are the *current* dollar holdings. For the purposes of optimization, as above, it is more convenient to deal with the holding weights w_i instead of the dollar holdings D_i . Let $\tilde{L}_i \equiv I L_i$, $D_i \equiv I w_i$ and $D_i^* \equiv I w_i^*$. Then

$$\tilde{P} = \frac{P}{I} = \sum_{i=1}^N R_i w_i - \sum_{i=1}^N L_i |w_i - w_i^*| \quad (133)$$

As above, we can have constraints

$$\sum_{i=1}^N w_i Y_{ia} = 0, \quad a = 1, \dots, m \quad (134)$$

We will assume that the current holdings satisfy the same constraints:

$$\sum_{i=1}^N w_i^* Y_{ia} = 0, \quad a = 1, \dots, m \quad (135)$$

This includes establishing trades ($w_i^* \equiv 0$). We have the normalization condition (55) for w_i , but not necessarily for w_i^* (*e.g.*, if the position is being established).

¹⁵ For $\rho = \pm 1$ the Sharpe ratio goes to infinity if $R_A \neq \pm R_B$: the volatility vanishes for $D_B = \mp D_A$ (recall that $\sigma_A = \sigma_B$). If $R_A = \pm R_B$, then the two instruments A and $-$ long for plus sign and short for minus sign $-B$ are indistinguishable for optimization purposes.

¹⁶ For the sake of simplicity, the transaction costs for buys and sells are assumed to be the same.

5.2 Optimization with Costs and Homogeneous Constraints

More generally, costs can be modeled by some function $f(w)$ of w_i , which also depends on the current holding weights w_i^* , but the precise form of this dependence is not going to be important here. We have

$$\tilde{P} = \sum_{i=1}^N R_i w_i - f \quad (136)$$

Generally, the costs spoil the invariance of the Sharpe ratio

$$S = \frac{\tilde{P}}{\tilde{V}} = \frac{\sum_{i=1}^N R_i w_i - f}{\sqrt{\sum_{i=1}^N C_{ij} w_i w_j}} \quad (137)$$

under the rescaling $w_i \rightarrow \zeta w_i$ ($\zeta > 0$) with a single exception of $f(w) = f_{\text{special}}(w)$:

$$f_{\text{special}}(w) \equiv \sum_{i=1}^N L_i |w_i| \quad (138)$$

where L_i are positive constants. The costs are of this form when we have only linear costs and the current holdings are zero, *i.e.*, this is an establishing trade. Below we will assume that $f(w)$ is *not* of this form.

In the absence of the rescaling invariance, care is needed when rewriting the Sharpe ratio maximization problem in terms of minimizing an objective function. The Sharpe ratio maximization problem reads:

$$\tilde{S} \equiv S + \sum_{a=1}^m \sum_{i=1}^N w_i Y_{ia} \mu_a + \tilde{\mu} \left(\sum_{i=1}^N |w_i| - 1 \right) \quad (139)$$

We need to maximize \tilde{S} w.r.t. w_i and Lagrange multipliers μ_a and $\tilde{\mu}$, which gives:¹⁷

$$\frac{1}{\tilde{V}} \left[R_i - f_i - \lambda \sum_{j=1}^N C_{ij} w_j \right] + \sum_{a=1}^m Y_{ia} \mu_a + \tilde{\mu} \text{sign}(w_i) = 0 \quad (140)$$

$$\sum_{i=1}^N w_i Y_{ia} = 0 \quad (141)$$

$$\sum_{i=1}^N |w_i| = 1 \quad (142)$$

$$\lambda \equiv \frac{\sum_{i=1}^N R_i w_i - f}{\sum_{i=1}^N C_{ij} w_i w_j} \quad (143)$$

$$f_i \equiv \frac{\partial f}{\partial w_i} \quad (144)$$

¹⁷ Actually, w_i derivatives are defined only for $w_i \neq 0$ and, *e.g.*, in the case of linear costs for $w_i \neq w_i^*$ – see Subsection 6.2 for details.

If we multiply the first equation by w_i and sum over i , we get

$$\tilde{\mu} = \frac{1}{\tilde{V}} \left[\sum_{i=1}^N f_i w_i - f \right] \quad (145)$$

Unless $f(w)$ has the special form (138), which we assume not to be the case, then generally $\tilde{\mu} \neq 0$.

We can still *formally* recast the Sharpe ratio maximization in terms of minimizing the following objective function (w.r.t. w_i and Lagrange multipliers μ'_a and $\tilde{\mu}'$):

$$g(w, \mu', \tilde{\mu}', \lambda') \equiv \frac{\lambda'}{2} \sum_{i,j=1}^N C_{ij} w_i w_j - \sum_{i=1}^N R_i w_i + f - \sum_{a=1}^m \sum_{i=1}^N w_i Y_{ia} \mu'_a - \tilde{\mu}' \left(\sum_{i=1}^N |w_i| - 1 \right) \quad (146)$$

$$g(w, \mu', \tilde{\mu}', \lambda') \rightarrow \min \quad (147)$$

The minimization equations read:

$$\lambda' \sum_{j=1}^N C_{ij} w_j - R_i + f_i - \sum_{a=1}^m Y_{ia} \mu'_a - \tilde{\mu}' \text{sign}(w_i) = 0 \quad (148)$$

$$\sum_{i=1}^N w_i Y_{ia} = 0 \quad (149)$$

$$\sum_{i=1}^N |w_i| = 1 \quad (150)$$

Multiplying the first equation by w_i and summing over i , we get

$$\tilde{\mu}' = \lambda' \sum_{i,j=1}^N C_{ij} w_i w_j - \sum_{i=1}^N [R_i - f_i] w_i \quad (151)$$

The statement then is that there exists a value of λ' for which minimizing the objective function produces the same solution for w_i as maximizing the Sharpe ratio. This value is given by $\lambda' = \lambda$, where λ is given by (143) with w_i corresponding to the *optimal solution*, *i.e.*, the maximal Sharpe ratio solution. We then have

$$\lambda' = \lambda \quad (152)$$

$$\mu'_a = \tilde{V} \mu_a \quad (153)$$

$$\tilde{\mu}' = \tilde{V} \tilde{\mu} = \sum_{i=1}^N f_i w_i - f \quad (154)$$

However, the practical value of this statement is limited – unless we solve the Sharpe ratio maximization problem, we do not know what λ is. The Sharpe ratio maximization problem is highly nonlinear and prone to usual nonlinear instabilities. On the other hand, in terms of minimizing the objective function, we can treat λ' as a parameter. Then the problem of maximizing the Sharpe ratio reduces to a one-dimensional problem of finding the value of λ' for which the Sharpe ratio is maximal.

5.2.1 Pitfalls

Because in the presence of costs the rescaling invariance is lost, the maximum Sharpe ratio solution is *not* given by the following minimization (w.r.t. w_i and Lagrange multipliers μ''_a):

$$\begin{aligned} \tilde{g}(w, \mu'', \lambda'') &\equiv \frac{\lambda''}{2} \sum_{i,j=1}^N C_{ij} w_i w_j - \sum_{i=1}^N R_i w_i + f - \\ &\quad - \sum_{a=1}^m \sum_{i=1}^N w_i Y_{ia} \mu''_a \end{aligned} \quad (155)$$

$$\tilde{g}(w, \mu'', \lambda'') \rightarrow \min \quad (156)$$

It is incorrect to assume – and this appears to be a common misstep in practical applications – that the maximum Sharpe ratio solution is given by the solution to this minimization condition for the value of λ'' such that (142) is satisfied.

To see this, for the sake of simplicity, let us assume that there are no linear constraints (141). Then we have

$$\lambda'' \sum_{j=1}^N C_{ij} w_j - R_i + f_i = 0 \quad (157)$$

Multiplying this equation by w_i and summing over i , we get

$$\lambda'' = \frac{\sum_{i=1}^N [R_i - f_i] w_i}{\sum_{i,j=1}^N C_{ij} w_i w_j} \quad (158)$$

Then, for this solution to coincide with (148) (without homogeneous constraints (141), that is), we must have

$$\frac{1}{\sum_{i,j=1}^N C_{ij} w_i w_j} \sum_{j=1}^N C_{ij} w_j = \text{sign}(w_i) \quad (159)$$

where we have taken into account that $\tilde{\mu}' \neq 0$ – see (154). Plugging (159) back into (157), we get

$$R_i - f_i = \gamma \text{sign}(w_i) \quad (160)$$

$$\gamma \equiv \sum_{i=1}^N [R_i - f_i] w_i \quad (161)$$

However, this is impossible to satisfy for a general form of $f(w)$. *E.g.*, in the case of linear costs

$$f(w) = f_{\text{linear}}(w) \equiv \sum_{i=1}^N L_i |w_i - w_i^*| \quad (162)$$

and we cannot have $R_i - L_i \text{sign}(w_i - w_i^*) = \gamma \text{sign}(w_i)$ for all $i = 1, \dots, N$. Therefore, minimizing the objective function (155) does *not* produce the maximal Sharpe ratio solution. The correct objective function to minimize is (146), where λ' is treated as a parameter, whose value is fixed via a one-dimensional search algorithm such that the Sharpe ratio is maximized.

5.2.2 Global vs. Local Optima

Assuming the cost function $f(w)$ is convex, the “wrong” objective function (155) is convex w.r.t. w_i , so it has a unique local minimum. However, the correct objective function (146) is not necessarily convex. This is because the contribution due to the $\tilde{\mu}'$ term is convex if and only if $\tilde{\mu}' \leq 0$, which is not necessarily the case – see (154). If $\tilde{\mu}' > 0$, there can be multiple local minima further complicating the search for a global minimum. *E.g.*, in the case of linear costs (162), we have $\tilde{\mu}' = \sum_{i=1}^N L_i w_i^* \text{sign}(w_i - w_i^*)$, which need not be negative.

5.3 Maximizing Sharpe Ratio with Linear Costs

As we saw in the previous subsection, in the presence of costs the Sharpe ratio maximization problem is highly nonlinear and may not even have a unique local minimum – even for linear costs (162), assuming some $w_i^* \neq 0$.

So, how is this optimization done in practice? Often it is done by simply taking the “wrong” objective function (155) and iterating λ'' until (55) is satisfied.¹⁸ As discussed above, this solution does not maximize the Sharpe ratio. In some cases, it could be a reasonable approximation though. Let us focus on linear costs. Let: i) L_i be uniform, $L_i \equiv L$; ii) $w_i^* \rightarrow w_i$ be a rebalancing trade from a previously optimized solution with $\sum_{i=1}^N |w_i^*| = 1$; iii) our portfolio be dollar neutral, so that $\sum_{i=1}^N w_i^* = \sum_{i=1}^N w_i = 0$; iv) the number of stocks be large ($N \gtrsim 1000$); and v) there be diversification constraints in place, so w_i are not larger than, say, low single digit percent.¹⁹ If $\text{sign}(w_i^*)$ and $\text{sign}(w_i - w_i^*)$ (*i.e.*, current holding signs and desired trade signs) are not highly correlated, then $|\tilde{\mu}'| \ll L$ and its contribution to (148) is small compared with the contribution due to the linear costs, so we can approximately ignore it. And neglecting the $\tilde{\mu}'$ contribution in (146) is the same as using (155).²⁰

With the above in mind, we can minimize the objective function (155) and fix λ'' via an iterative procedure until (55) is satisfied. This is what is done in most

¹⁸ For any $\lambda'' > 0$, there is a unique optimum assuming C_{ij} is positive-definite, and all $L_i \geq 0$.

¹⁹ We will discuss bounds below. Alternatively, one can squash returns to achieve the same.

²⁰ This argument also goes through for partially establishing and liquidating trades with $\xi \lesssim 1$, *i.e.*, $\xi \equiv \sum_{i=1}^N |w_i^*|$ need not be equal 1. Uniformity of L_i can also be relaxed (with some care).

practical applications. This also avoids the issue of multiple local optima discussed in Subsection 5.2.2 – the objective function (155) is convex and has a unique local minimum. However, we emphasize: this is only an *approximation* to Sharpe \rightarrow max.

6 Optimization: Costs, Constraints & Bounds

So, let us consider the following optimization problem:

$$g(w, \mu, \lambda) \equiv \frac{\lambda}{2} \sum_{i,j=1}^N C_{ij} w_i w_j - \sum_{i=1}^N (R_i w_i - L_i |w_i - w_i^*|) - \sum_{a=1}^m \sum_{i=1}^N w_i Y_{ia} \mu_a \quad (163)$$

$$g(w, \mu, \lambda) \rightarrow \min \quad (164)$$

$$w_i^- \leq w_i \leq w_i^+ \quad (165)$$

where: the minimization is w.r.t. w_i and Lagrange multipliers μ_a ; λ is treated as a parameter to be fixed iteratively so that the normalization condition

$$\sum_{i=1}^N |w_i| = 1 \quad (166)$$

is satisfied; and we have included lower w_i^- and upper w_i^+ bounds (165) on the holding weights. If there are no bounds, we can simply take w_i^- and w_i^+ to be large negative and large positive numbers, respectively.

In the following it will be more convenient to use

$$x_i \equiv w_i - w_i^* \quad (167)$$

We then have

$$\tilde{g}(x, \mu, \lambda) \equiv \frac{\lambda}{2} \sum_{i,j=1}^N C_{ij} x_i x_j - \sum_{i=1}^N (\rho_i x_i - L_i |x_i|) - \sum_{a=1}^m \sum_{i=1}^N x_i Y_{ia} \mu_a \quad (168)$$

$$\tilde{g}(x, \mu, \lambda) \rightarrow \min \quad (169)$$

$$x_i^- \leq x_i \leq x_i^+ \quad (170)$$

$$\rho_i \equiv R_i - \lambda \sum_{j=1}^N C_{ij} w_j^* \quad (171)$$

$$x_i^\pm \equiv w_i^\pm - w_i^* \quad (172)$$

Furthermore, we are assuming that the current holdings satisfy the linear constraints

$$\sum_{i=1}^N w_i^* Y_{ia} = 0, \quad a = 1, \dots, m \quad (173)$$

and we have dropped an immaterial constant term from the objective function (168).

6.1 Bounds

Typically, in practical applications, bounds are used to cap i) the positions of individual stocks in a portfolio, and ii) the amount of trading in each stock. *E.g.*, let us assume that we impose the following constraints:

$$|w_i| \leq \xi \tilde{v}_i \quad (174)$$

$$|w_i - w_i^*| \leq \tilde{\xi} \tilde{v}_i \quad (175)$$

$$\tilde{v}_i \equiv \frac{v_i}{I} \quad (176)$$

where v_i is, say, a 20-day average daily dollar volume for the stock labeled by i , and ξ and $\tilde{\xi}$ are some positive percentages. Then we would have

$$x_i^+ = \min \left(\tilde{\xi} \tilde{v}_i, \xi \tilde{v}_i - w_i^* \right) \geq 0 \quad (177)$$

$$x_i^- = \max \left(-\tilde{\xi} \tilde{v}_i, -\xi \tilde{v}_i - w_i^* \right) \leq 0 \quad (178)$$

and we are assuming that $|w_i^*| \leq \xi \tilde{v}_i$. There is little to no value in trying to account for any isolated “extraordinary” cases (*e.g.*, there is news for a given stock and it needs to be liquidated, which for $w_i^* > 0$ would mean that $x_i^+ < 0$, and for $w_i^* < 0$ it would mean that $x_i^- > 0$) as they can be simply treated by setting the desired holdings for such few stocks and altogether excluding them from optimization of the remaining universe of stocks. We will therefore assume that $x_i^+ \geq 0$ and $x_i^- \leq 0$.

We can avoid much notational headache if we further assume that $x_i^+ > 0$ and $x_i^- < 0$. There are cases where one may wish to set $x_i^+ = 0$ or $x_i^- = 0$, *e.g.*, we cannot sell a stock due to a short-sale restriction (hard-to-borrow stock, *etc.*). However, instead of setting x_i^+ or x_i^- strictly to zero, it is more practical to set it to a small positive or negative number instead (*e.g.*, within the desired precision tolerance). In the following we will assume that $x_i^+ > 0$ and $x_i^- < 0$.

6.2 Optimization: General Case

Let us define the following subsets of the index $i = 1, \dots, N$:

$$x_i \neq 0, \quad i \in J \quad (179)$$

$$x_i = 0, \quad i \in J' \quad (180)$$

$$x_i = x_i^+ > 0, \quad i \in J^+ \subset J \quad (181)$$

$$x_i = x_i^- < 0, \quad i \in J^- \subset J \quad (182)$$

$$\bar{J} \equiv J^+ \cup J^- \quad (183)$$

$$\tilde{J} \equiv J \setminus \bar{J} \quad (184)$$

$$\eta_i \equiv \text{sign}(x_i), \quad i \in J \quad (185)$$

Note that, since the modulus has a discontinuous derivative, the minimization equations are not the same as setting first derivatives of $\tilde{g}(x, \mu, \lambda)$ w.r.t. x_i and μ_a to

zero. More concretely, first derivatives w.r.t. x_i are well-defined for $i \in J$, but not for $i \in J'$, while the first derivatives w.r.t. μ_a are always well-defined. Furthermore, first derivatives w.r.t. x_i for $i \in \tilde{J}$ (i.e., at the bounds) need not be zero. Let us therefore consider the global minimum condition:

$$\tilde{g}(x', \mu', \lambda)|_{x'_i=x_i+\epsilon_i, \mu'_a=\mu_a+\epsilon_a} \geq \tilde{g}(x_i, \mu_a, \lambda) \quad (186)$$

Here ϵ_i and ϵ_a *a priori* are arbitrary except that at the bounds we have

$$\epsilon_i \leq 0, \quad i \in J^+ \quad (187)$$

$$\epsilon_i \geq 0, \quad i \in J^- \quad (188)$$

From (186) we get

$$\begin{aligned} & \frac{\lambda}{2} \sum_{i,j=1}^N C_{ij} \epsilon_i \epsilon_j + \sum_{i \in J} L_i (|x_i + \epsilon_i| - |x_i| - \eta_i \epsilon_i) - \sum_{a=1}^m \sum_{i=1}^N \epsilon_i Y_{ia} \epsilon_a - \\ & \sum_{a=1}^m \sum_{i=1}^N x_i Y_{ia} \epsilon_a + \sum_{j=1}^N \left(\lambda \sum_{i \in J} C_{ij} x_i - \rho_j + L_j \tilde{\eta}_j - \sum_{a=1}^m Y_{ja} \mu_a \right) \epsilon_j \geq 0 \end{aligned} \quad (189)$$

where (the ambiguity in $\text{sign}(\epsilon_i)$ below is immaterial; we can set $\text{sign}(0) = 0$)

$$\tilde{\eta}_i \equiv \eta_i, \quad i \in J \quad (190)$$

$$\tilde{\eta}_i \equiv \text{sign}(\epsilon_i), \quad i \in J' \quad (191)$$

The first line in (189) is $\mathcal{O}(\epsilon^2)$. For infinitesimal ϵ_i the second line gives:

$$\lambda \sum_{j \in J} C_{ij} x_j - \rho_i + L_i \eta_i - \sum_{a=1}^m Y_{ia} \mu_a = 0, \quad i \in \tilde{J} \quad (192)$$

$$\sum_{i=1}^N x_i Y_{ia} = \sum_{i \in J} x_i Y_{ia} = 0, \quad a = 1, \dots, m \quad (193)$$

which equations correspond to setting to zero first derivatives of $\tilde{g}(x, \mu, \lambda)$ w.r.t. x_i , $i \in \tilde{J}$ and μ_a , and then we also have the following inequalities for $i \notin \tilde{J}$:

$$\forall j \in J' : \quad \left| \lambda \sum_{i \in J} C_{ij} x_i - \rho_j - \sum_{a=1}^m Y_{ja} \mu_a \right| \leq L_j \quad (194)$$

$$\forall j \in J^+ : \quad \lambda \sum_{i \in J} C_{ij} x_i - \rho_j - \sum_{a=1}^m Y_{ja} \mu_a \leq -L_j \quad (195)$$

$$\forall j \in J^- : \quad \lambda \sum_{i \in J} C_{ij} x_i - \rho_j - \sum_{a=1}^m Y_{ja} \mu_a \geq L_j \quad (196)$$

With (192), (193), (194), (195) and (196), the second line in (189) is positive-definite for all ϵ_i (subject to (187) and (188), that is) – this is because these terms are linear in ϵ_i . On the other hand, the first term in the first line of (189) is positive semi-definite as C_{ij} is assumed to be positive-definite. The second term is positive semi-definite as $\eta_i = \text{sign}(x_i)$ for $i \in J$. The third term implies that for any $\epsilon_a \neq 0$ we have the following condition on ϵ_i :

$$\sum_{i=1}^N \epsilon_i Y_{ia} = 0 \quad (197)$$

This is simply the condition that we are only allowed to consider paths $x_i \rightarrow x_i + \epsilon_i$ along which the constraints (193) are satisfied (for all $i \in \{1, \dots, N\}$).

The conditions (194), (195) and (196) must be satisfied by the solution to (192) and (193), which give a global optimum. However, even ignoring the bounds for a moment, *a priori* we do not know i) what the subset J' is and ii) what the values of η_i are for $i \in J$, so we have 3^N – a prohibitively large number – possible combinations.

6.3 Optimization: Factor Model

This can be circumvented by assuming the factor model form for C_{ij} :

$$C_{ij} = \Theta_{ij} \equiv \xi_i^2 \delta_{ij} + \sum_{A=1}^K \Omega_{iA} \Omega_{jA} \quad (198)$$

Here, any values of A such that the corresponding column of Ω_{iA} is a linear combination of the columns of Y_{ia} must be omitted (with the specific risk untouched). This is because in (192), (194), (195) and (196) C_{ij} appears only in the combination

$$\sum_{j \in J} C_{ij} x_j = \xi_i^2 x_i + \sum_{A=1}^K \Omega_{iA} \sum_{j \in J} x_j \Omega_{jA}, \quad i \in J \quad (199)$$

$$\sum_{j \in J} C_{ij} x_j = \sum_{A=1}^K \Omega_{iA} \sum_{j \in J} x_j \Omega_{jA}, \quad i \in J' \quad (200)$$

so if any column in Ω_{iA} is a linear combination of the columns of Y_{ia} , its contribution vanishes due to (193). We assume that all such columns in Ω_{iA} , if any, are omitted.

The optimization problem reduces to solving a $(K+m)$ -dimensional system. Let

$$v_A \equiv \sum_{i=1}^N x_i \Omega_{iA} = \sum_{i \in J} x_i \Omega_{iA}, \quad A = 1, \dots, K \quad (201)$$

Further, let $H \equiv \{a\} \cup \{A\}$. Let $\widehat{\Omega}_{i\alpha}$, $\alpha \in H$ be the following $N \times (K+m)$ matrix

$$\widehat{\Omega}_{ia} \equiv Y_{ia} \quad (202)$$

$$\widehat{\Omega}_{iA} \equiv \Omega_{iA} \quad (203)$$

Let u_α be the following $(K + M)$ -vector:

$$u_a \equiv -\frac{1}{\lambda} \mu_a \quad (204)$$

$$u_A \equiv v_A \quad (205)$$

From (192), (193) and (201) we have

$$x_i = \frac{1}{\lambda \xi_i^2} \left(\rho_i - L_i \eta_i - \lambda \sum_{\alpha \in H} \hat{\Omega}_{i\alpha} u_\alpha \right), \quad i \in \tilde{J} \quad (206)$$

$$\sum_{i \in J} x_i \hat{\Omega}_{i\alpha} = \sum_{\beta \in H} \varphi_{\alpha\beta} u_\beta \quad (207)$$

where $\varphi_{\alpha\beta}$ is the following symmetric $(K + m) \times (K + m)$ matrix:

$$\varphi_{AB} \equiv \delta_{AB} \quad (208)$$

$$\varphi_{Ab} = 0 \quad (209)$$

$$\varphi_{ab} = 0 \quad (210)$$

Recalling that we have

$$x_i \eta_i > 0, \quad i \in \tilde{J} \quad (211)$$

we get

$$\eta_i = \text{sign} \left(\rho_i - \lambda \sum_{\alpha \in H} \hat{\Omega}_{i\alpha} u_\alpha \right), \quad i \in \tilde{J} \quad (212)$$

$$\forall i \in J^+ : \quad \rho_i - \lambda \sum_{\alpha \in H} \hat{\Omega}_{i\alpha} u_\alpha \geq L_i + \lambda \xi_i^2 x_i^+ \equiv L_i^+ \quad (213)$$

$$\forall i \in J^- : \quad \rho_i - \lambda \sum_{\alpha \in H} \hat{\Omega}_{i\alpha} u_\alpha \leq -L_i + \lambda \xi_i^2 x_i^- \equiv -L_i^- \quad (214)$$

$$\forall i \in \tilde{J} : \quad \left| \rho_i - \lambda \sum_{\alpha \in H} \hat{\Omega}_{i\alpha} u_\alpha \right| > L_i \quad (215)$$

$$\forall i \in J' : \quad \left| \rho_i - \lambda \sum_{\alpha \in H} \hat{\Omega}_{i\alpha} u_\alpha \right| \leq L_i \quad (216)$$

where (215) follows from (211) and (206). The last four inequalities define J^+ , J^- , \tilde{J} and J' in terms of $(K + m)$ unknowns u_α . Note that $L_i^\pm > L_i$, $i \in J^\pm$ and if we take $x_i^\pm \rightarrow \pm\infty$, we get empty J^\pm .

Substituting (206) into (207), we get the following system of $(K + m)$ equations for $(K + m)$ unknowns u_α :

$$\sum_{\beta \in H} \hat{Q}_{\alpha\beta} u_\beta = y_\alpha \quad (217)$$

where

$$\widehat{Q}_{\alpha\beta} \equiv \varphi_{\alpha\beta} + \sum_{i \in \widetilde{J}} \frac{\widehat{\Omega}_{i\alpha} \widehat{\Omega}_{i\beta}}{\xi_i^2} \quad (218)$$

$$y_\alpha \equiv \frac{1}{\lambda} \sum_{i \in \widetilde{J}} \frac{\widehat{\Omega}_{i\alpha}}{\xi_i^2} [\rho_i - L_i \eta_i] + \sum_{i \in J^+} x_i^+ \widehat{\Omega}_{i\alpha} + \sum_{i \in J^-} x_i^- \widehat{\Omega}_{i\alpha} \quad (219)$$

so we have

$$u_\alpha = \sum_{\beta \in H} \widehat{Q}_{\alpha\beta}^{-1} y_\beta \quad (220)$$

where \widehat{Q}^{-1} is the inverse of \widehat{Q} .

Note that (220) solves for u_α given η_i , J^+ , J^- , \widetilde{J} and J' . On the other hand, (212), (213), (214), (215) and (216) determine η_i , J^+ , J^- , \widetilde{J} and J' in terms of u_α . The entire system is then solved iteratively, where at the initial iteration one takes $\widetilde{J}^{(0)} = \{1, \dots, N\}$, so that $J^{+(0)}$, $J^{-(0)}$ and $J'^{(0)}$ are empty, and

$$\eta_i^{(0)} = \pm 1, \quad i = 1, \dots, N \quad (221)$$

While *a priori* the values of $\eta_i^{(0)}$ can be arbitrary, unless $(K + m) \ll N$, in some cases one might encounter convergence speed issues. However, if one chooses

$$\eta_i^{(0)} = \text{sign}(\rho_i), \quad i = 1, \dots, N \quad (222)$$

then the iterative procedure generally is expected to converge rather fast.

The following trick can speed up the convergence. Let $\widehat{x}_i^{(s)}$ be such that

$$\forall i \in \{1, \dots, N\} : \quad x_i^- \leq \widehat{x}_i^{(s)} \leq x_i^+ \quad (223)$$

$$\sum_{i=1}^N \widehat{x}_i^{(s)} Y_{ia} = 0, \quad a = 1, \dots, m \quad (224)$$

Let $x_i^{(s+1)}$ be the solution obtained at the $(s+1)$ -th iteration. This solution satisfies the linear constraints, but may not satisfy the bounds. Let

$$q_i \equiv x_i^{(s+1)} - \widehat{x}_i^{(s)} \quad (225)$$

$$h_i(t) \equiv \widehat{x}_i^{(s)} + t q_i, \quad t \in [0, 1] \quad (226)$$

Then

$$\widehat{x}_i^{(s+1)} \equiv h_i(t_*) = \widehat{x}_i^{(s)} + t_* q_i \quad (227)$$

where t_* is the maximal value of t such that $h_i(t)$ satisfies the bounds. We have:

$$q_i > 0 : \quad p_i \equiv \min \left(x_i^{(s+1)}, x_i^+ \right) \quad (228)$$

$$q_i < 0 : \quad p_i \equiv \max \left(x_i^{(s+1)}, x_i^- \right) \quad (229)$$

$$t_* = \min \left(\frac{p_i - \widehat{x}_i^{(s)}}{q_i} \mid q_i \neq 0, i = 1, \dots, N \right) \quad (230)$$

Now, at each step, instead of (213) and (214), we can define J^+ and J^- via (J' is still defined via (216))

$$\forall i \in J^+ : \quad \hat{x}_i = x_i^+ \quad (231)$$

$$\forall i \in J^- : \quad \hat{x}_i = x_i^- \quad (232)$$

where \hat{x}_i is computed iteratively as above and we can take $\hat{x}_i^{(0)} \equiv 0$ at the initial iteration. The difference between (231), (232) and (213), (214) is that the former add new elements to the sets J^+ and J^- one (or a few) element(s) at each iteration, while the latter can add many elements.

The convergence criteria are given by (this produces the global optimum)

$$\tilde{J}^{(s+1)} = \tilde{J}^{(s)} \quad (233)$$

$$J^{+(s+1)} = J^{+(s)} \quad (234)$$

$$J^{-(s+1)} = J^{-(s)} \quad (235)$$

$$\forall i \in \tilde{J}^{(s+1)} : \quad \eta_i^{(s+1)} = \eta_i^{(s)} \quad (236)$$

$$\forall \alpha \in H : \quad u_\alpha^{(s+1)} = u_\alpha^{(s)} \quad (237)$$

The first four of these criteria are based on discrete quantities and are unaffected by computational (machine) precision effects, while the last criterion is based on continuous quantities and in practice is understood as satisfied within computational (machine) precision or preset tolerance.

7 Example: Intraday Mean-Reversion Alpha

In this section, to illustrate our discussion in Section 2, we discuss an intraday mean-reversion alpha. Let us set up our notations. P_i , $i = 1, \dots, N$ is the stock price for the stock labeled by i , where N is the number of stocks in our universe. In actuality, the price for each stock is a time-series: P_{is} , $s = 0, 1, \dots, M$, where the index s labels trading dates, with $s = 0$ corresponding to the most recent date in the time series. We will use superscript O (unadjusted open price), C (unadjusted close price), AO (open price fully adjusted for splits and dividends), and AC (close price fully adjusted for splits and dividends), so, *e.g.*, P_{is}^C is the unadjusted close price. V_{is} is the unadjusted daily volume (in shares, not dollars). We define the overnight return as the close-to-next-open return:

$$R_{is} \equiv \ln (P_{is}^{AO} / P_{i,s+1}^{AC}) \quad (238)$$

Note that both prices in this definition are fully adjusted.

Next, we take an $N \times K$ binary loadings matrix Λ_{iA} for our universe in three incarnations, based on Bloomberg Industry Classification System (BICS) sectors, industries and sub-industries. These are binary clusters discussed in Subsection

2.4.²¹ For each date s , we cross-sectionally regress our returns R_{is} over Λ_{iA} with no intercept²² and unit weights, as in (23). We take the residuals ε_{is} of the regression (23), and specify the desired dollar holdings via

$$D_{is} = -\varepsilon_{is} \frac{I}{\sum_{j=1}^N |\varepsilon_{js}|} \quad (239)$$

$$\sum_{i=1}^N |D_{is}| = I \quad (240)$$

$$\sum_{i=1}^N D_{is} = 0 \quad (241)$$

where I is the *intraday* investment level, which is the same for all dates s .

The portfolio is established at the open²³ assuming fills at the open prices P_{is}^O , and liquidated at the close on the same day assuming fills at the close prices P_{is}^C , with no transaction costs or slippage, both of which are present in real life – here our goal is not to build a realistic trading strategy that will make money in real life, but to illustrate our discussion in Section 2. Daily P&L for each stock is given by

$$\Pi_{is} = D_{is} \left[\frac{P_{is}^C}{P_{is}^O} - 1 \right] \quad (242)$$

The shares bought plus sold (*i.e.*, for the establishing and liquidating trades combined) for each stock on each day are computed via $Q_{is} = 2|D_{is}|/P_{is}^O$.

Before we can run our regressions, we need to select our universe. We wish to keep our discussion here as simple as possible, so we select our universe based on the average daily dollar volume (ADDV) defined via

$$A_{is} \equiv \frac{1}{d} \sum_{r=1}^d V_{i,s+r} P_{i,s+r}^C \quad (243)$$

We take $d = 21$ (*i.e.*, one month), and then take our universe to be top 2000 tickers by ADDV. However, to ensure that we do not inadvertently introduce a universe selection bias,²⁴ we do not rebalance the universe daily. Instead, we rebalance monthly, every 21 trading days, to be precise. *I.e.*, we break our 5-year backtest period (see below) into 21-day intervals, we compute the universe using ADDV (which,

²¹ Note that stocks rarely jump (sub-)industries/sectors, so Λ_{iA} can be assumed to be static.

²² More precisely, the intercept is already subsumed in Λ_{iA} : $\Lambda_{iA} = 1$ if the stock labeled by i belongs to the cluster labeled by $A = 1, \dots, K$; otherwise, $\Lambda_{iA} = 0$. Each stock belongs to one and only one cluster. This implies that $\sum_{A=1}^K \Lambda_{iA} = 1$ for each i , so a linear combination of the columns of Λ_{iA} is the intercept.

²³ This is a so-called “delay-0” alpha – P_{is}^O is used in the alpha, and as the establishing fill price.

²⁴ *I.e.*, to ensure that our results are not a mere consequence of the universe selection.

in turn, is computed based on the 21-day period immediately preceding such interval), and use this universe during the entire such interval.²⁵ The bias that we do have, however, is the survivorship bias. We take the data for the universe of tickers as of 9/6/2014 that have historical pricing data on <http://finance.yahoo.com> (accessed on 9/6/2014) for the period 8/1/2008 through 9/5/2014. We restrict this universe to include only U.S. listed common stocks and class shares (no OTCs, preferred shares, *etc.*) with BICS sector, industry and sub-industry assignments as of 9/6/2014.²⁶ However, it does not appear that the survivorship bias is a leading effect here (see below). Also, ADDV-based universe selection is by no means optimal and is chosen here for the sake of simplicity. In practical applications, the trading universe of liquid stocks is carefully selected based on market cap, liquidity (ADDV), price and other (proprietary) criteria.

We run our simulation over a period of 5 years. More precisely, $M = 252 \times 5$, and $s = 0$ is 9/5/2014 (see above). The results for the annualized return-on-capital (ROC), annualized Sharpe ratio (SR) and cents-per-share (CPS) are given in Table 1 for 3 choices of clusters: BICS sectors, industries and sub-industries. ROC is computed as average daily P&L divided by the investment level I (with no leverage) and multiplied by 252. SR is computed as daily Sharpe ratio multiplied by $\sqrt{252}$. CPS is computed as the total P&L divided by total shares traded. The P&L graphs for the 3 cases in Table 1 are given in Figure 1.

In the above model we have done no risk management apart from (automatic in this case) dollar neutrality. We can do risk management via weighted regression as in Subsection 2.7. However, here we will discuss another method. The basic issue is that some residuals ϵ_{is} can be very large, so the strategy can disproportionately load up on stocks with such large residuals and as a result the portfolio is not diversified enough. This is why SR in Table 1 are not as high as in Table 2 (see below). We can deal with such large residuals by treating them as outliers. One well-known method is Winsorization. Here we discuss a conceptually similar method, which is more convenient. Let X_i be a set of values for which we expect to have a normal distribution with cross-sectional mean \bar{X} and standard deviation χ . Let \tilde{X}_i be the values of X_i deformed such that \tilde{X}_i are conformed to the normal distribution with the same mean \bar{X} and standard deviation χ . *E.g.*, we can use the `normalize()` function given in Appendix A of (Kakushadze and Liew, 2014). Now let us apply this method to our residuals ϵ_{is} separately for each date (so everything is out-of-sample). Let the resulting values be $\tilde{\epsilon}_{is}$. Note that $\tilde{\epsilon}_{is}$ still have vanishing cross-sectional means, but the outliers have been “squashed”. We can now use $\tilde{\epsilon}_{is}$ instead of ϵ_{is} in (239) and maintain dollar neutrality. The results for ROC, SR and CPS are given in Table 2. Note a dramatic increase in SR compared with Table 1 – at the expense of lowering

²⁵ Note that, since the alpha is purely intraday, this “rebalancing” does not generate additional trades, it simply changes the universe that is traded for the next 21 days.

²⁶ The number of such tickers in our data is 3,811. The number of BICS sectors is 10. The numbers of BICS industries is 48. The number of BICS sub-industries varies between 164 and 169 (due to small sub-industries, which are affected by the varying top-2000-by-ADDV universe).

ROC and CPS. The P&L graphs for the 3 cases in Table 2 are given in Figure 2.

One evident caveat of this alpha is that the “open” can be a fuzzy notion as some stocks do not always open at 9:30:00 sharp. So, our assumption that the orders can be placed simultaneously at the open is a bit faulty.²⁷ In real life one would have to wait until a little after the open and compute the alpha for the stocks that are open as of that time, then send the orders and get fills. An intraday simulated strategy of this type can be accessed freely on <http://vynance.com/portfolio.html>. The establishing time is 9:31:30 (and the liquidating time is 15:59:00). The trading universe varies from day-to-day and is smaller than 2000 tickers, it is roughly in the range of 200-300 tickers for long positions and 200-300 tickers for short positions. The performance for this strategy from 2/18/2014 through 9/19/2014 is as follows: ROC = 29.19%; SR = 12.13; CPS = 1.52. The simulation assumes no trading costs or slippage; however, it is not a “delay-0” but (more realistic) “delay-30-seconds” strategy, *i.e.*, the alpha used for the establishing trades at 9:31:30 is computed based on the pricing data from 9:31:00 (which data itself is delayed 5-35 seconds). Since “inception” (4/14/2011) Vynance Portfolio has had a consistent simulated performance with the annualized daily Sharpe ratio of about 16 and monthly return-on-capital of about 3%. This is consistent with our results in Table 2, which indicates that the survivorship bias in our results should not be a leading effect – Vynance Portfolio simulations are done daily, in real time, and thus are free from the survivorship bias.

8 Concluding Remarks

As mentioned earlier, there are a myriad ways of doing mean-reversion. In the quantitative framework we discussed in these notes, a mean-reversion model is essentially defined by the risk factors used as loadings in regressions along with regression weights, or, in optimization, by the choice of the multi-factor risk model and constraints – the latter usually also relating to neutrality w.r.t. some risk factors.

Here we should emphasize that there are all kinds of bells and whistles one can add to tweak a particular mean-reversion model, even within the aforementioned framework. Also, if only regressions are used, then factor covariance matrix is not needed and one can settle for using, *e.g.*, historical volatilities in regression weights, *i.e.*, in this case one only needs the unrotated factor loadings matrix $\tilde{\Omega}_{iA}$. On the other hand, for optimization a full multi-factor risk model is required – not just factor loadings matrix, but also the factor covariance matrix and specific risk. Since depending on the choice of the constraints – and also the returns used – one may omit some risk factors from the factor loadings matrix, in many cases it is warranted to compute a custom multi-factor risk model. This topic is covered in much more detail in (Kakushadze and Liew, 2014) dedicated to this subject.

²⁷ Plus we are assuming “delay-0”, meaning, we can place the trades “infinitely” fast right after receiving the opening prints from the exchange(s) and get filled at the very same open prices. And, as mentioned above, we are ignoring trading costs and slippage, hence the rosy ROC, SR and CPS.

References

- Adcock, J.C. and Meade, N. (1994) A simple algorithm to incorporate transactions costs in quadratic optimization. *European Journal of Operational Research* 79(1): 85-94.
- Allaj, E. (2013) The Black-Litterman Model: A Consistent Estimation of the Parameter Tau. *Financial Markets and Portfolio Management* 27(2): 217-251.
- Atkinson, C., Pliska, S.R. and Wilmott, P. (1997) Portfolio management with transaction costs. *Proc. Roy. Soc. London Ser. A* 453(1958): 551-562.
- Avellaneda, M. and Lee, J.H. (2010) Statistical arbitrage in the U.S. equity market. *Quant. Finance* 10(7): 761-782.
- Best, M.J. and Hlouskova, J. (2003) Portfolio selection and transactions costs. *Computational Optimization and Applications* 24(1): 95-116.
- Black, F. and Litterman, R. (1991) Asset allocation: Combining investors views with market equilibrium. *J. Fixed Income* 1(2): 7-18.
- Black, F. and Litterman, R. (1992) Global portfolio optimization. *Financ. Anal. J.* 1992, 48(5): 28-43.
- Cadenillas, A. and Pliska, S.R. (1999) Optimal trading of a security when there are taxes and transaction costs. *Finance and Stochastics* 3(2): 137-165.
- Cheung, W. (2010) The Black-Litterman model explained. *Journal of Asset Management* 11(4): 229-243.
- Chu, B., Knight, J. and Satchell, S.E. (2011) Large Deviations Theorems for Optimal Investment Problems with Large Portfolios. *European Journal of Operations Research* 211(3): 533-555.
- Cvitanović, J. and Karatzas, I. (1996) Hedging and portfolio optimization under transaction costs: a martingale approach. *Math. Finance* 6(2): 133-165.
- Daniel, K. (2001) The Power and Size of Mean Reversion Tests. *Journal of Empirical Finance* 8(5): 493-535.
- Da Silva, A.S., Lee, W. and Pornrojngangkool, B. (2009) The Black-Litterman model for active portfolio management. *Journal of Portfolio Management* 35(2): 61-70.
- Davis, M. and Norman, A. (1990) Portfolio selection with transaction costs. *Math. Oper. Res.* 15(4): 676-713.

- Drobetz, W. (2001) How to Avoid the Pitfalls in Portfolio Optimization? Putting the Black-Litterman Approach at Work. *Financ. Mark. Portf. Manag.* 15(1): 59-75.
- Dumas, B. and Luciano, E. (1991) An exact solution to a dynamic portfolio choice problem under transaction costs. *The Journal of Finance* 46(2): 577-595.
- Fama, E.F. and MacBeth, J.D. (1973) Risk, Return and Equilibrium: Empirical Tests. *J. Polit. Econ.* 81(3): 607-636.
- Fama, E.F. and French, K.R. (1993) Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* 33(1): 3-56.
- Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G. (2006) Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *Review of Financial Studies* 19(3): 797-827.
- He, G. and Litterman, R. (1999) The Intuition Behind Black-Litterman Model Portfolio. San Francisco, CA: Goldman Sachs and Co, Investment Management Division.
- Hodges, S. and Carverhill, A. (1993) Quasi mean reversion in an efficient stock market: the characterization of economic equilibria which support Black-Scholes option pricing. *The Economic Journal* 103(417): 395-405.
- Idzorek, T. (2007) A Step-by-Step Guide to the Black-Litterman Model. In: Satchell, S. (ed.) *Forecasting Expected Returns in the Financial Markets*. Waltham, MA: Academic Press.
- Janeček, K. and Shreve, S. (2004) Asymptotic analysis for optimal investment and consumption with transaction costs. *Finance Stoch.* 8(2), 181-206.
- Jegadeesh, N. and Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance* 48(1): 65-91.
- Jegadeesh, N. and Titman, S. (1995) Overreaction, delayed reaction, and contrarian profits. *Rev. Financ. Stud.* 8(4): 973-993.
- Liew, J. and Roberts, R. (2013) U.S. Equity Mean-Reversion Examined. *Risks* 1(3): 162-175.
- Kakushadze, Z. and Liew, J. (2014) Custom v. Standardized Risk Models. SSRN Working Paper, <http://ssrn.com/abstract=2493379> (September 8, 2014); arXiv:1409.2575.
- Knight, J. and Satchell, S.E. (2010) Exact properties of measures of optimal investment for benchmarked portfolios. *Quantitative Finance* 10(5): 495-502.

- Lo, W.A. and MacKinlay, A.C. (1990) When are contrarian profits due to stock market overreaction? *Rev. Financ. Stud.* 3(2): 175-205.
- Lo, A.W. (2010) *Hedge Funds: An Analytic Perspective*. Princeton University Press, p. 260.
- Magill, M. and Constantinides, G. (1976) Portfolio selection with transactions costs. *J. Econom. Theory* 13(2): 245-263.
- Markowitz, H. (1952) Portfolio selection. *Journal of Finance* 7(1): 77-91.
- Merton, R.C. (1969) Lifetime portfolio selection under uncertainty: the continuous time case. *The Review of Economics and Statistics* 51(3): 247-257.
- Mitchell, J.E. and Braun, S. (2013) Rebalancing an investment portfolio in the presence of convex transaction costs, including market impact costs. *Optimization Methods and Software* 28(3): 523-542.
- Mokkhavesa, S. and Atkinson, C. (2002) Perturbation solution of optimal portfolio theory with transaction costs for any utility function. *IMA J. Manag. Math.* 13(2): 131-151.
- O'Tool, R. (2013) The Black-Litterman model: A risk budgeting perspective. *Journal of Asset Management* 14(1): 2-13.
- Perold, A.F. (1984) Large-scale portfolio optimization. *Management Science* 30(10): 1143-1160.
- Poterba, M.J. and Summers, L.H. (1988) Mean reversion in stock prices: Evidence and implications. *J. Financ. Econ.* 1988, 22(1): 27-59.
- Rockafellar, R.T. and Uryasev, S. (2000) Optimization of conditional value-at-risk. *Journal of Risk* 2(3): 21-41.
- Satchell, S. and Scowcroft, A. (2000) A demystification of the Black-Litterman model: Managing quantitative and traditional portfolio construction. *Journal of Asset Management* 1(2): 138-150.
- Sharpe, W.F. (1966) Mutual fund performance. *Journal of Business* 39(1): 119-138.
- Shreve, S. and Soner, H.M. (1994) Optimal investment and consumption with transaction costs. *Ann. Appl. Probab.* 4(3): 609-692.

Table 1: Simulation results for the intraday mean-reversion alphas discussed in Section 7 without normalizing the regression residuals.

Clusters	ROC	SR	CPS
BICS Sectors	44.58%	6.21	1.17
BICS Industries	49.00%	7.15	1.29
BICS Sub-industries	51.77%	7.87	1.36

Table 2: Simulation results for the intraday mean-reversion alphas discussed in Section 7 with normalizing the regression residuals.

Clusters	ROC	SR	CPS
BICS Sectors	33.27%	11.55	1.02
BICS Industries	37.67%	15.02	1.15
BICS Sub-industries	40.40%	18.50	1.24

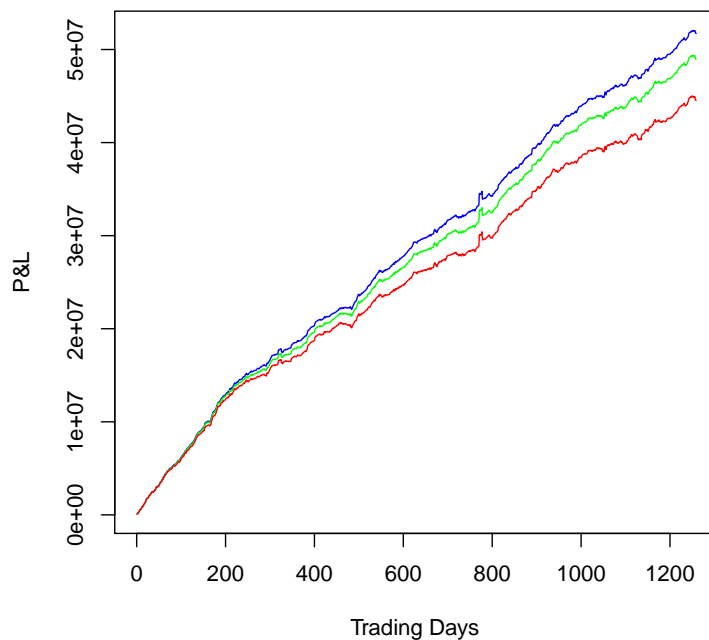


Figure 1. P&L graphs for the mean-reversion alpha (unnormalized residuals) discussed in Section 7, with a summary in Table 1. Bottom-to-top-performing: i) BICS sectors, ii) BICS industries, and iii) BICS sub-industries. The investment level is \$10M long plus \$10M short.

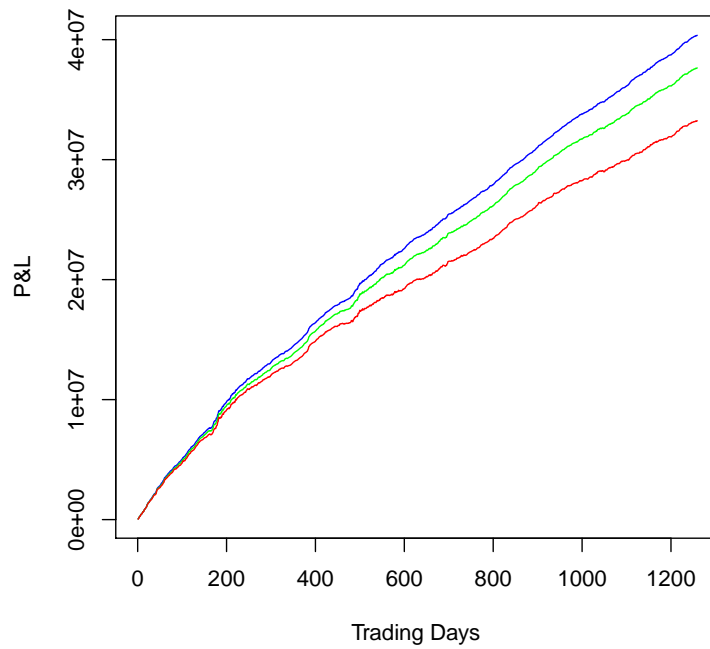


Figure 2. P&L graphs for the mean-reversion alpha (normalized residuals) discussed in Section 7, with a summary in Table 2. Bottom-to-top-performing: i) BICS sectors, ii) BICS industries, and iii) BICS sub-industries. The investment level is \$10M long plus \$10M short.