Portfolio Optimization using Factor Models

Abstract

We identify three possible scenarios when the optimization algorithm fails to produce a reasonable answer, and show the adjustments that must be made to the optimization input parameters in each case in order to guarantee a well behaved resulting portfolio

Portfolio optimization is an algorithm that for a given level of expected P&L, produces a portfolio with the lowest volatility. Mathematically, optimizing a portfolio entails maximizing the risk-adjusted P&L given by the formula: $\alpha w^T - \frac{1}{2}\lambda w \Sigma w^T$ where α is a row vector ($1 \times n$) of expected returns, w is a row vector ($1 \times n$) of dollar investments in each asset, λ determines the importance assigned to risk vs. expected P&L, and Σ is an $n \times n$ co-variance matrix of asset returns. The expression $\alpha w^T - \frac{1}{2}\lambda w \Sigma w^T$ achieves maximum value when $w = \frac{1}{\lambda}\alpha \Sigma^{-1}$

The naive approach to generating Σ is to compute historical co-variances between all pairs of assets. Unfortunately, not only is there a finite amount of price history, but also companies' characteristics and accordingly co-variances change over time. As a consequence, the amount of available data is very limited. In practice, the ratio of the number of assets to the number of historical data points is such that the co-variance matrix is either singular, or close enough to being singular that computing an optimal portfolio using it would be like running a regression with 1000 observations and 500 variables; a very high R^2 but virtually no out of sample predictive power whatsoever.

This problem is solved in practice by using a factor model, namely, breaking down the co-variance matrix as follows: $\Sigma = \Lambda \phi \Lambda^T + \Psi$. As a result of using a factor model, the overall portfolio risk $w \Sigma w^T$ can be broken down into factor risk $w(\Lambda \phi \Lambda^T) w^T$ and specific risk $w \Psi w^T$.

Minimizing specific risk is straightforward. If we have no factor risk, then $\Sigma = \Psi$, so the optimal solution is going to be $w = \frac{1}{\lambda}\alpha \Psi^{-1}$, but since Ψ is a diagonal matrix, it's inverse will simply be the reciprocal of it's diagonal elements: $diag(a_1, \ldots, a_n)^{-1} = diag\left(\frac{1}{a_1}, \ldots, \frac{1}{a_n}\right)$. In other words, specific risk is minimized when the investment level in each stock is proportionate to it's expected return divided by it's specific variance, that is: $w_i \propto \frac{\alpha_i}{\psi_i}$. Mathematically, this creates a well diversified portfolio where the investment level is commensurate with the expected return, which is precisely what we want.

So under what conditions will this produce concentrated holdings rather than a well diversified portfolio? If all the alphas are zero (or close to zero) except for a handful of stocks, the portfolio won't be diversified. So most of the alphas need to be non-zero, and of similar magnitude. The only other way we could get concentrated holdings is if some of the specific risks are too close to zero. But so long as both of these conditions are checked for and avoided, a low-risk, diversified portfolio is guaranteed.

The situation with factor risk is a bit more complicated. If we have no specific risk, then $\Sigma=\Lambda\phi\Lambda^T$ so the optimal solution is going to be: $w=\frac{1}{\lambda}\alpha\left(\Lambda\phi\Lambda^T\right)^{-1}$. Initially, this looks problematic, because $\Lambda\phi\Lambda^T$ is singular. Luckily, as it turns out, $\alpha\left(\Lambda\phi\Lambda^T\right)^{-1}$ is equal to residuals of regressing alphas on factor loadings. In other words, while $\Lambda\phi\Lambda^T$ can't be inverted directly, if you add a small constant to the diagonal, it does become invertable and as this constant approaches zero, the solution approaches regression residuals (Appendix A). Intuitively, this makes sense because factor risk is minimized when there is no factor exposure (i.e. when $w\Lambda\sqrt{\phi}=0$), and regression residuals always satisfy this condition (Appendix B).

Those familiar with regression are used to the idea that a high R^2 is good. But in this case R^2 measures how much P&L we have to give up in order to minimize factor risk, so a high R^2 is very bad. However, even if the R^2 is 1, there's an easy way to pick how much P&L we want to give up in order to reduce factor risk. In order to produce a higher P&L, but higher factor risk portfolio, scale up the specific risk; that is, instead of using $\Sigma=\Lambda\phi\Lambda^T+\Psi$, use $\Sigma=\Lambda\phi\Lambda^T+m\Psi$ where m > 1. If you want a lower factor risk, but lower P&L portfolio, make m < 1.

There's a caveat to this though. Say you use a reasonably good factor model that captures most of the correlation between assets to construct a well diversified portfolio with little or no factor risk. The probability distribution of that portfolio's P&L will be normal, via the central limit theorem, so variance will fully and accurately describe it's risk. This is not the case for portfolios with lots of factor risk, such as long only portfolios; it is well known that their returns are anything but normally distributed, making variance a flawed and potentially inappropriate risk measure in this case.

Appendix A

Prove: $resid(lm(\alpha \sim -1+\Lambda)) = \lim_{m\to\infty} \frac{1}{m}\alpha\left(\Lambda\phi\Lambda^T + \frac{1}{m}I\right)^{-1}$

where

α is a row vector ($1\times n$)

Λ is an $n\times k$ matrix

ϕ is a $k\times k$ matrix

m is a single value

Regression coefficients are given by the formula $\beta=(X^T X)^{-1}X^T y$ where y is a column vector. Regression fitted values, and consequently residuals, don't depend on rotation of the xs; in other words $resid(lm(\alpha\sim -1+\Lambda)) = resid(lm(\alpha\sim -1+\Lambda\sqrt{\phi}))$. Mathematically, let z be any square invertable matrix. Regressing a variable y on matrix X will produce fitted values equal to: $X\beta = X(X^T X)^{-1}X^T y$. The fitted values of regressing y on X multiplied by z only requires us to replace X with (X z) in the formula above as follows:

$$(X z)((X z)^T (X z))^{-1}(X z)^T y = X z (z^T X^T X z)^{-1} z^T X^T y = X z z^{-1} (X^T X)^{-1} (z^T)^{-1} z^T X^T y =$$
$$X I (X^T X)^{-1} I X^T y = X (X^T X)^{-1} X^T y = X\beta$$

Therefore, if we let $L=\Lambda\sqrt{\phi}$, it is sufficient to prove that $(\alpha^T - L\beta)^T = \lim_{m\to\infty}\frac{1}{m}\alpha\left(L L^T + \frac{1}{m}I\right)^{-1}$

Transposing both sides of the equation we get: $\alpha^T - L\beta = \lim_{m\to\infty}\frac{1}{m}\left[\left(L L^T+\frac{1}{m}I\right)^{-1}\right]^T \alpha^T$. Because the $L L^T + \frac{1}{m} I$ matrix is symmetrical, so is its inverse, and we need to prove that $\alpha^T - L\beta = \lim_{m\to\infty}\frac{1}{m}\left(L L^T + \frac{1}{m}I\right)^{-1}\alpha^T$ where $\beta = (L^T L)^{-1} L^T \alpha^T$

Start with the Woodbury matrix identity, $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+V A^{-1} U)^{-1} V A^{-1}$, and plug the following values into it: $A = I_{n\times n}$, $U = L\sqrt{m}$, $C = I_{k\times k}$, and $V = (L\sqrt{m})^T$

The left side of the identity becomes:

$$(A+UCV)^{-1} = \left(I_{n\times n} + (L\sqrt{m}) I_{k\times k} (L\sqrt{m})^T\right)^{-1} = (I + m L L^T)^{-1} = \left[m\left(\frac{1}{m}I + L L^T\right)\right]^{-1} = \frac{1}{m}\left(L L^T + \frac{1}{m}I\right)^{-1}$$

The right side becomes:

$$A^{-1} - A^{-1}U(C^{-1} + V A^{-1} U)^{-1} V A^{-1} =$$
$$I_{n\times n}{}^{-1} - I_{n\times n}{}^{-1} L\sqrt{m}(I_{k\times k}{}^{-1} + (L\sqrt{m})^T I_{n\times n}{}^{-1} L\sqrt{m})^{-1} (L\sqrt{m})^T I_{n\times n}{}^{-1} =$$
$$I - (L\sqrt{m})(I_{k\times k} + (L\sqrt{m})^T (L\sqrt{m}))^{-1} (L\sqrt{m})^T = \text{ (factoring out } \sqrt{m}\text{)}$$
$$I - m L\left(I_{k\times k} + m L^T L\right)^{-1} L^T = I - m L\left(m[\frac{1}{m} I_{k\times k} + L^T L]\right)^{-1} L^T = I - m L \frac{1}{m}\left(\frac{1}{m} I_{k\times k} + L^T L\right)^{-1} L^T =$$

$$\boxed{I - L\left(\frac{1}{m} I_{k\times k} + L^T L\right)^{-1} L^T}$$

We combine the left and right sides of the identity,

$$\frac{1}{m}\left(L L^T+\frac{1}{m}I\right)^{-1}=I-L\left(\frac{1}{m}I_{k\times k}+L^T L\right)^{-1}L^T$$

and multiply both sides of the resulting equation by α^T :

$$\frac{1}{m}\left(L L^T+\frac{1}{m}I\right)^{-1}\alpha^T=\left(I-L\left(\frac{1}{m}I_{k\times k}+L^T L\right)^{-1}L^T\right)\alpha^T=\alpha^T-L\left(\frac{1}{m}I_{k\times k}+L^T L\right)^{-1}L^T\alpha^T$$

Using the above equation, we can re-write the limit we're trying to prove as follows:

$$\lim_{m\to\infty}\frac{1}{m}\left(L L^T+\frac{1}{m}I\right)^{-1}\alpha^T=\lim_{m\to\infty}\left[\alpha^T-L\left(\frac{1}{m}I_{k\times k}+L^T L\right)^{-1}L^T\alpha^T\right]=\alpha^T-L\lim_{m\to\infty}\left[\left(\frac{1}{m}I_{k\times k}+L^T L\right)^{-1}\right]L^T\alpha^T$$

Now, $\left(\frac{1}{m}I_{k\times k}+L^T L\right)^{-1}$ is a continuous function of m, because each element of the inverse (row i, column j) will be the determinant of (i,j) minor (which is just a sum of products, no discontinuity there) divided by determinant of $\frac{1}{m}I_{k\times k}+L^T L$ and $\left|\frac{1}{m}I_{k\times k}+L^T L\right|\geq\left|\frac{1}{m}I_{k\times k}\right|+|L^T L|$, so no discontinuity there either, therefore:

$$\lim_{m\to\infty}\left(\frac{1}{m}I_{k\times k}+L^T L\right)^{-1}=(L^T L)^{-1}$$

and therefore

$$\alpha^T-L\lim_{m\to\infty}\left[\left(\frac{1}{m}I_{k\times k}+L^T L\right)^{-1}\right]L^T\alpha^T=\alpha^T-L(L^T L)^{-1}L^T\alpha^T$$

Substituting $\beta=(L^T L)^{-1}L^T\alpha^T$ into the equation above, we see that indeed

$$\lim_{m\to\infty}\frac{1}{m}\left(L L^T+\frac{1}{m}I\right)^{-1}\alpha^T=\alpha^T-L\beta$$

Appendix B

Recall that $w\Lambda\sqrt{\phi}=wL=(L^T w^T)^T$ and given that $w^T=\alpha^T-L(L^T L)^{-1}L^T\alpha^T$,

$$L^T w^T=L^T(\alpha^T-L(L^T L)^{-1}L^T\alpha^T)=L^T\alpha^T-L^T L(L^T L)^{-1}L^T\alpha^T=L^T\alpha^T-I L^T\alpha^T=0$$

In English, it means that one of the properties of regression residuals is that when they are multiplied by X's, the result is always a vector of zeros, in other words, regression residuals always produce a zero factor risk portfolio (i.e. satisfy the condition $w\Lambda\sqrt{\phi}=0$)