6. Multi-Factor Models and Optimization

6.1. **Portfolio Risk Analysis.** We begin by reviewing the multi-factor models we introduced last time. These models assume a linear functional form

$$R_{t+1} = X_t f_{t+1} + \epsilon_{t+1}, \quad \mathbb{E}[\epsilon] = 0, \quad \mathbb{V}[\epsilon] = D \tag{6.1}$$

where R_{t+1} is an n-dimensional random vector containing the cross-section of returns in excess of the risk-free rate over some time interval $[t, t+1]$, and X_t is a (non-random) $n \times p$ matrix that can be calculated entirely from data known before time t. As before, when we are doing cross-sectional analysis for fixed t, we will drop the explicit time-subscripts, for example writing (6.1) as $R = Xf + \epsilon$.

Also ϵ_{t+1} is assumed to follow a mean-zero distribution with diagonal variance-covariance matrix

$$D := \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2) \ \text{ with all } \ \sigma_i^2 > 0. \tag{6.2}$$

The variable f in (6.1) denotes a p-dimensional random vector process which cannot be observed directly; information about the f-process must be obtained via statistical inference. We assume that the f-process is stationary with finite first and second moments:

$$\mathbb{E}[f] = \mu_f, \quad \text{and} \quad \mathbb{V}[f] = F. \tag{6.3}$$

We also recall the definition of "exposure of the portfolio" to one or more factors:

Definition 6.1. For a portfolio with holdings vector $\boldsymbol{h} \in \mathbb{R}^n$, the vector

$$X'h \in \mathbb{R}^p$$

is called the *exposure vector* of the portfolio. We will use this often, so we introduce the shorthand notation $x := X'h$. The j-th element of x, denoted x_j, is called the *exposure* of h to the j-th factor.

The model (6.1), (6.2) and (6.3) entails associated reductions of the first and second moments of the asset returns:

$$\mathbb{E}[R] = X\mu_f, \quad \text{and} \quad \Sigma := \mathbb{V}[R] = D + XFX' \tag{6.4}$$

where X' denotes the transpose. Eq. (6.4) is quite useful for portfolio construction and for analyzing existing portfolios. For example, it says that

$$h'\Sigma h = h'Dh + h'XFX'h = h'Dh + x'Fx \tag{6.5}$$

which expresses the portfolio's variance in terms of the *idiosyncratic variance* $h'Dh$ and a second term computable from only the exposure vector.

We can transform (6.5) into an equivalent form which makes the relative contributions of different terms easier to interpret. Divide both sides by the total variance, $h'\Sigma h$ to get

$$1 = \underbrace{\frac{h'Dh}{h'\Sigma h}}_{\text{idiosyncratic}} + \sum_{i=1}^{p} \underbrace{x_i \sum_j \frac{F_{ij} x_j}{h'\Sigma h}}_{i\text{-th factor contribution}} \tag{6.6}$$

Eq. (6.6) is known as a *variance decomposition*. It has a term for the idiosyncratic variance, plus one term for each factor. Note that the idiosyncratic term is never negative, but the variance contribution from a factor can be negative if a factor is acting as a hedge and reducing variance caused by exposure to other factors.

The APT relation (6.1) allows us to attribute not only risk (variance), but also return or P&L. In particular, the portfolio's one-period P&L is given by

$$h'R = h'\epsilon + x'f = h'\epsilon + \sum_i x_i f_i\,. \tag{6.7}$$

Schematically, (6.7) means that

$$\text{PL} = \text{idiosyncratic} + \sum_i [\text{contribution from } i\text{-th factor}]$$

Eq.(6.7) is called a *performance attribution*.

The ability to decompose the profit and loss in this way can be very powerful. For example, one can determine whether a given portfolio manager's skill is greater in stock-selection (idiosyncratic) or in timing exposure to various factors or categories of factors. The manager may not be aware of the true sources of their returns. For example, a manager could be under the impression that they are doing stock selection, but their research process could be leading them into stocks with certain beta or momentum characteristics, and it could be that after controlling for these effects, their value-add is zero or negative even if they have outperformed the market over a certain period.

However, applying (6.7) in practice also carries with it significant challenges. One major challenge is that the term $h'R$ does not include slippage. I often refer to $h'R$ as *ideal PL*. It is ideal in the sense that it is what we could earn in an "ideal world" where we could execute an entire order at the midpoint price which prevailed at the time the order was constructed. This is a gross idealization because it ignores price moves during the period over which the order is executed, including those caused by our own impact on the price (ie. our market impact).

6.2. **Measures of Risk.** Three of the most common measures of risk are volatility, value-at-risk, and expected shortfall. Since you're already quite familiar with volatility, we now discuss the latter two.

Imagine you are an investment bank before the Volcker rule, and management gets nervous if the trading division loses more than \$50mm in a single day. If π is a random variable representing the profit in a day, then $\ell = -\pi$ is known as the *loss*. Suppose the strategists analyze the predictive density $p(\ell)$ and estimate that

$$\int_{5\times 10^7}^{\infty} p(\ell)d\ell \approx 0.01,$$

so the strategy will only make management nervous about once in every 100 days. In this situation, the number \$50mm equals the 99% VaR.

Supposing that

$$F_\ell(x) = \int_{-\infty}^{x} p(\ell)d\ell,$$

the c.d.f. of the loss distribution, is a one-to-one function and hence invertible, the 99% VaR is $F_\ell^{-1}(0.99)$. More generally, F_ℓ may not be invertible and there is nothing magical about the number 0.99. For any $\alpha \in (0,1)$, we define VaR_α to be

$$\mathrm{VaR}_\alpha = \inf\left\{x \in \mathbb{R} : \alpha \leq F_\ell(x)\right\}.$$

In terminology, VaR_α is called the "$100 \times (1-\alpha)$% VaR" so $\alpha = 0.05$ is 95% VaR etc. The statistically-minded will recognize this as the definition of the *quantile function* and is thus an old idea.

For the normal distribution, the VaR is a simple scaling of the volatility. The scale factor to use is determined by the quantile function of the standard normal. For example, here are a few typical values.

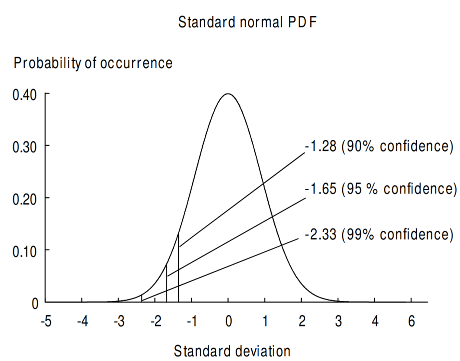

FIGURE 6.1. This chart shows three confidence level scaling factors and their associated tail probability of loss levels.

The problem with VaR is that it has nothing to say about the *size of the loss*, when it does occur. For illustrative purposes consider a hypothetical strategy whose profit distribution takes the form of a mixture of two distributions. The first is a normal distribution with daily sigma of \$1m and mean annual profit of \$80m (hence a Sharpe ratio of 5). The second is a distribution which is usually zero, but once in 1000 trading days (about four years) there is a "catastrophe" which causes the strategy to lose \$1,000m.

Suppose all risk-management were done by means of VaR and that all traders were compensated with an annual performance bonus, without too much visibility into what the individual traders were doing. Unscrupulous or nescient traders might be drawn to strategies of the type considered above. Moreover, it wouldn't be surprising to find such strategies readily available in the marketplace, since the expected gain of taking the other side or "selling the strategy" is \$680m every four years.

This example, although stylized, illustrates the problem with relying too heavily on VaR. For this reason, VaR has amusingly been compared by Einhorn and Brown (2008) to "an air bag which works all of the time, except when you have a car accident." Einhorn further charged that VaR:

(1) Led to excessive risk-taking and leverage at financial institutions
(2) Focused on the manageable risks near the center of the distribution and ignored the tails
(3) Created an incentive to take "excessive but remote risks"
(4) Was potentially catastrophic when its use creates a false sense of security among senior executives and watchdogs.

VaR isn't the ultimate panacea we might have hoped for.

Attempts have been made to formulate risk measures which address some of the shortcomings of VaR. One of the more promising ones is *expected shortfall*, defined as the conditional expected loss, conditional on the event that the loss is greater than the VaR. Expected shortfall is also called Conditional Value at Risk (CVaR), and expected tail loss (ETL).

Mathematically, if the underlying distribution for X is a continuous distribution then the expected shortfall is equivalent to the tail conditional expectation defined by

$$\mathrm{TCE}_\alpha(X) = \mathbb{E}[-X \mid X \leq \mathrm{VaR}_\alpha(X)].$$

This is a nice measure theoretically. The main problem is that it's very hard to measure empirically, because there are, by definition, fewer real historical events where $X \leq -\mathrm{VaR}_\alpha(X)$.

If the payoff of a portfolio X follows normal (Gaussian) distribution with the density $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ then the expected shortfall is equal to

$$-\mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}$$

where $\phi(x)$ is the density of the standard normal, and $\Phi(x)$ is the standard normal cumulative density, so $\Phi^{-1}(\alpha)$ is the standard normal quantile.

6.3. **Homogeneous Risk Measures.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *homogeneous of degree k* if

$$f(\lambda x) = \lambda^k f(x) \text{ for all } \lambda > 0, x \in \mathbb{R}^n.$$

Three of the most commonly-used risk measures are volatility, value-at-risk, and expected shortfall. They are all homogeneous of degree $k = 1$. For example, the volatility of a 2:1 levered portfolio is theoretically twice the volatility of the unlevered version. It is worth noting that, for a very large portfolio, if it had to be liquidated quickly then the losses could be magnified due to the market impact of the liquidation itself. This special case contradicts the homogeneity statement above, but nonetheless we shall continue to investigate the consequences of homogeneity.

Euler's homogeneous function theorem states that f is homogeneous of degree k if and only if

$$x \cdot \nabla f = \sum_i x_i \frac{\partial f}{\partial x_i} = k f(x)\,. \tag{6.8}$$

This is typically applied with $k = 1$, as follows.

Quite generally, say some return denoted r (which could be our portfolio's return $h'R$) can be written

$$r = \sum_m x_m g_m \tag{6.9}$$

where x_m are non-random "exposures" known at the beginning of the period (known *ex ante*, in other words) and g_m are random return sources whose realizations become known *ex post*, or at the end of the period.

Then by Euler's theorem (6.8) one can write

$$\sigma(r) = \sum_m x_m \,\mathrm{MCR}_m, \quad \text{where} \quad \mathrm{MCR}_m := \frac{\partial \sigma(r)}{\partial x_m} \tag{6.10}$$

where MCR is for *marginal contribution to risk*. Applying the definition of covariance one can also derive from (6.9) another variance decomposition,

$$\sigma^2(r) = \mathrm{cov}(r, r) = \sum_m x_m \,\mathrm{cov}(g_m, r)$$

Dividing the last equation by $\sigma(r)$ yields the *x-sigma-rho attribution*

$$\sigma(r) = \sum_m x_m \sigma(g_m) \rho(g_m, r) \tag{6.11}$$

where $\rho(g_m, r)$ is the correlation of source m with the portfolio's return. By comparing (6.11) with (6.10) we see that

$$\mathrm{MCR}_m = \sigma(g_m)\rho(g_m, r)$$

Theorem 6.1. For an unconstrained optimal portfolio (i.e., maximum information ratio), the expected source returns are directly proportional to the source marginal contributions,

$$\mathbb{E}[g_m] = \mathrm{IR} \cdot \mathrm{MCR}_m$$

where IR is the portfolio information ratio.

Proof. If we are at optimal $\mathrm{IR} = \mathbb{E}[r]/\sigma(r)$, then

$$0 = \frac{\partial}{\partial x_i}\left[\frac{\mathbb{E}(r)}{\sigma(r)}\right] = \frac{\sigma(r)\mathbb{E}[g_i] - \mathbb{E}[r]\,\mathrm{MCR}_i}{\sigma(r)^2}$$

Hence $\sigma(r)\mathbb{E}[g_i] = \mathbb{E}[r]\,\mathrm{MCR}_i$, from which we get the desired relation by dividing by $\sigma(r)$ on both sides. $\square$

An easy way to remember Theorem 6.1 is to say: at optimality, marginal contributions to risk are proportional to marginal contributions to return.

Theorem 6.1 provides implied returns that serve as an important reality check on whether the actual portfolio is consistent with the manager's views. In any situation where we can calculate MCR's, we can then calculate $\mathrm{IR}{\cdot}\mathrm{MCR}_m$ for a few reasonable choices of IR. Sometimes non-quantitative portfolio managers are surprised by the results. The directions and signs line up to make sense: if $\mathrm{MCR}_m < 0$ for some m, it means you could reduce risk by having more exposure to it. This would happen, for example, if source m were uncorrelated to the other sources, and you were short (negative exposure), which would happen if $\mathbb{E}[g_m] < 0$.

6.4. **Optimization and APT.** The Markowitz (1952) mean-variance problem with moments (6.4) is

$$h^* = \operatorname{argmax} f(h) \quad \text{where} \tag{6.12}$$

$$f(h) = h'X\mu_f - \frac{\kappa}{2}h'XFX'h - \frac{\kappa}{2}h'Dh \tag{6.13}$$

and where $\kappa > 0$ is the Arrow-Pratt constant absolute risk aversion.

The first two terms in (6.12) depend on h only through its exposures $x := X'h$. In terms of x the first two terms in (6.12) can be written more simply as

$$x'\mu_f - (\kappa/2)x'Fx$$

It turns out that at optimality, the third term can be written as a function of x as well:

Intuition 6.1. An optimal portfolio h for (6.12) must minimize idiosyncratic variance $h'Dh$ among all portfolios with the same exposure vector $x := X'h$.

With this intuition in mind to clarify the proof, we can now proceed to the main result.

Theorem 6.2. The risk/alpha exposures of the portfolio optimizing (6.12) are x^* and the optimal holdings are h^*, where

$$\begin{aligned} x^* &= \kappa^{-1}[F + [X'D^{-1}X]^+]^{-1}\mu_f \\ h^* &= \kappa^{-1}D^{-1/2}(X'D^{-1/2})^+\left[F + (X'D^{-1}X)^+\right]^{-1}\mu_f \end{aligned} \tag{6.14}$$

Proof. Let $h^*(x)$ be the solution to

$$h^*(x) = \operatorname*{argmin}_h h'Dh \quad \text{subject to:} \quad X'h = x. \tag{6.15}$$

Let

$$V(x) = h^*(x)'Dh^*(x)$$

be the minimum idiosyncratic variance (still subject to $X'h = x$).

Using Intuition 6.1, the mean-variance objective can then be written entirely in terms of x:

$$\max_x \left\{ x \cdot \mu_f - \frac{\kappa}{2}x'Fx - \frac{\kappa}{2}V(x) \right\}. \tag{6.16}$$

Finding $V(x)$ will allow us to directly attack (6.16), hence we now devote ourselves to this task. We show in due course that $V(x)$ is quadratic and can be written down explicitly.

Changing variables to $\eta := D^{1/2}h$ the problem (6.15) is

$$\min_\eta \|\eta\|^2 \text{ subject to } X'D^{-1/2}\eta = x \tag{6.17}$$

Hence the solution to (6.17) is given by[1]

$$\eta^* = (X'D^{-1/2})^+x \quad \Rightarrow \quad h^*(x) = D^{-1/2}(X'D^{-1/2})^+x \tag{6.18}$$

and therefore

$$\begin{aligned} V(x) &= h^*(x)'Dh^*(x) \\ &= [D^{-1/2}(X'D^{-1/2})^+x]'D[D^{-1/2}(X'D^{-1/2})^+x] \\ &= x'[X'D^{-1}X]^+x \end{aligned}$$

The third term in (6.16) can thus be combined with the second term in (6.16) to form a single quadratic term. The optimal x is then given by

$$x^* = \kappa^{-1}[F + [X'D^{-1}X]^+]^{-1}\mu_f$$

and the optimal holdings h^* are found by plugging x^* into (6.18). This completes the proof. Theorem 6.2 and further discussion and examples can be found in: http://ssrn.com/abstract=2821360.

[1]Under certain conditions on matrices A and B, one has $(AB)^+ = B^+A^+$ but none of those conditions apply here, so the right-hand side of (6.18) can't be simplified further.

It is not necessary to assume that X is of full rank to apply Eqns. (6.14). Suppose for simplicity that $D = \sigma^2 I$ for some constant $\sigma^2 > 0$. Then in the formula

$$x^* = \kappa^{-1}[F + [X'X]^+]^{-1}\mu_f$$

we would be computing the Moore-Penrose pseudoinverse $[X'X]^+$ which exists for any matrix X, whether or not it is full rank. Also, one can show easily that $[X'X]^+$ is positive semi-definite, hence the matrix $F + [X'X]^+$ will be invertible as long as F is invertible. The latter assumption – invertibility of F – is the only real assumption here.

This method therefore deals gracefully with approximate or exact colinearities among the risk factors and alpha factors. Such colinearities could arise in any of the following cases: (a) X contains indicator variables for two classifications, such as sector and country, or (b) if two or more alpha models were closely related representations of the same model/dataset, or (c) if some group of alpha factors were approximately spanned by the risk factors. Eqns. (6.14) remain valid if there *aren't* any collinearities of course, so this approach allows one formula to cover all cases.

Theorem 6.3. Let the number of factors, p, be a fixed constant. The computational complexity of finding the optimal exposures and holdings, Eqns. (6.14) is linear-time in n, the number of assets.

Proof. Recall that if $X = USV'$ is the SVD of X, the Moore-Penrose pseudoinverse is given by

$$X^+ = VS^+U' \tag{6.19}$$

where S^+ is formed by replacing every non-zero diagonal entry by its reciprocal and transposing the resulting matrix. Let $p \ll n$; the "economical" SVD of an $n \times p$ matrix can be computed (Golub and Van Loan, 2012) in about

$$6np^2 + 20p^3 \text{ flops} \tag{6.20}$$

This bounds the complexity of the pseudoinverse and actual inverse required in (6.14), since the Moore-Penrose pseudoinverse is given in terms of the SVD by (6.19). But if p is constant then (6.20) is linear in n. This completes the proof.

The number of risk factors k, the number of alpha models m, and hence the overall number of factors $p = k + m$ is ultimately a modeling choice, but since the complexity scales as p^3 for fixed n, parsimonious models are more efficiently optimized. Parsimonious models are also preferred in statistical model selection procedures – see the Ockham's razor principle (Jefferys and Berger, 1992).

The technique above can be extended to include certain simple trading cost models. For example, if we have a starting portfolio h_0 and quadratic trading costs[2] given by

$$(h - h_0)'\Lambda(h - h_0) \quad \text{where} \quad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

then (6.12) becomes (with the shorthand $\alpha := X\mu_f$)

$$\begin{aligned} f(h) &= h'\alpha - \frac{\kappa}{2}h'XFX'h - \frac{\kappa}{2}h'Dh - (h - h_0)'\Lambda(h - h_0) \\ &= h'(\alpha + 2\Lambda h_0) - \frac{\kappa}{2}h'XFX'h - \frac{\kappa}{2}h'(D + \frac{2}{\kappa}\Lambda)h \end{aligned}$$

The latter has the same mathematical structure as the original problem (6.12), so Theorem 6.2 applies.

We lose no generality in assuming that the outputs from m distinct alpha models are stored in the first m columns in X, and defining $k = p - m$ as the number of risk factors, one has

$$X = \begin{bmatrix} X_\alpha & X_\sigma \end{bmatrix} \in \mathbb{R}^{n \times (k+m)} \tag{6.21}$$

The remarks below refer to the notation of (6.21).

If the alpha factors are statistically independent from the risk factors, then F must have a block structure with blocks F_α and F_σ. A portfolio *neutral* to all of the columns of X_σ (the risk factors), may be obtained as the limit of h^* as $[F_\sigma]_{i,i} \to +\infty$ for all $i = 1, \ldots, k$. This limit exists, and is equivalent to solving the optimization with the k linear constraints $h'X_\sigma = 0$. Under the independence assumption of alpha factors and risk factors, the constrained factor-neutral portfolio will usually be close to h^*, but not exactly the same since the constrained solution may have higher idiosyncratic variance than other unconstrained solutions.

References

Einhorn, David and Aaron Brown (2008). "Private profits and socialized risk". In: *Global Association of Risk Professionals* 42, pp. 10–26.

Golub, Gene H and Charles F Van Loan (2012). *Matrix computations.* Vol. 3. JHU Press.

Jefferys, William H and James O Berger (1992). "Ockham's razor and Bayesian analysis". In: *American Scientist* 80.1, pp. 64–72.

Markowitz, Harry (1952). "Portfolio selection*". In: *The Journal of Finance* 7.1, pp. 77–91.

[2] This trading cost model is too simple to be used in practice because the quadratic structure tends to underestimate the cost of small trades.