BASIC MATHEMATICS OF PORTFOLIO MANAGEMENT

.

1. PORTFOLIOS

Only the most reckless of investors will put every penny into the same investment. There are theoretical and practical advantages to diversifying into multiple investments.

While carrying multiple investments complicates the mathematics of finance tremendously, it it more than worth it. From a mathematical point of view, this is relatively obvious even from a distribution-free (or almost distribution-free) point of view.

Consider a disastrous scenario of total loss. Ignoring leverage, any single security S_1 may have some probability p_1 of going to zero. If we create a portfolio with both security S_1 and S_2 however, then the probability of disaster is equal to

$$P(S_1 \to 0 \ \& \ S_2 \to 0) = P(S_2 \to 0|S_1 \to 0)P(S_1 \to 0)$$

which of course is bounded above by $p(S_1 \to 0)$. Obviously the situation becomes even more favorable as more and more securities are added to the mix.

When securities are completely uncorrelated, the mathematics of portfolios becomes almost as simple as for single securities, since every quantity is separable. You might think in practice this ends up being so far from true as to be worthless speculation, but we will see that clever portfolio construction, as with factor neutrality, can bring about simple calculations.

1.1. **Inventory: the Ineluctable Investment Issue.** Any serious market investor, quantitative or otherwise, ends up buying and selling an array of securities subject to the willingness of other market participants to trade them. Since every type of investor is ultimately limited by capital constraints, the amount of inventory that can be held is finite. This leads to constraints on new investments, based on what is already in their portfolio.

1.2. **Notation.** We say here that investable securities are indexed over $i = 1, \ldots, N$. When considering returns, we use the letter

$$\boldsymbol{r} = (r_1, \ldots, r_N)^*$$

Expected returns are written

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^*$$

Where appropriate, the covariance matrix of returns is

$$\boldsymbol{\Sigma} = (\sigma_{ij})$$

Much portfolio theory considers investments over just a single period. This is a very useful simplifying assumption, but completely ignores the inventory problem.

We typically assume that a portfolio is constructed from a set of assets

$$\boldsymbol{S} = (S_1, \ldots, S_N)^*$$

with dollar investment of w_i in each asset, defining an investment vector

$$\boldsymbol{w} = (w_1, \ldots, w_N)^*$$

Another ridiculous, but useful, assumption is that the investor does not affect the markets when trading, meaning that an arbitrary amount w of any asset can be bought or sold[1]. If that's the case, then the total investment size doesn't matter. Rather, only relative sizes of the w have any effect. Therefore, it is normal to assume that

$$\sum_{i=1}^{N} w_i = 1$$

We will occasionally separate out cash with the index $i = 0$ but usually we will just leave it implicitly among the N available assets.

Our total returns, if the individual asset returns $\boldsymbol{r}$ are known, are given by the dot product

$$R = \boldsymbol{w}^* \boldsymbol{r} = \sum_{i=1}^{N} r_i w_i$$

2. Motives in Portfolio Management

2.1. **Mutual Funds.** A *mutual fund* invests money on behalf of its investors, charging fees as a proportion of investments. US mutual

---

[1]In practice, sophisticated investors in products that trade electronically tend to use "algos" provided by their broker in order to achieve an acceptable tradeoff between market impact and traded quantity. Investors in over-the-counter (OTC) products carefully cultivate personal relationships with an array of counter parties.

funds collectively hold well over $10T in assets. They are subject to fairly heavy regulation, and typically have tight restrictions on their investment universe.

As a result, mutual funds tend to be constrained in their portfolio choices to a degree that is very significant for portfolio management purposes.

Mutual fund customers typically evaluate them on

- the basis of their taste for the mutual fund's "style"
- their long-term returns
- subjective judgments of their volatility, and
- (all-too-seldom) management fees.

The motives for established mutual fund managers are therefore to achieve a tolerable return in their class, without losing too many existing customers. Up-and-coming managers are motivated to take more chances in order to attract attention and investors.

2.1.1. *Full Investment.* Mutual funds are often expected to be *fully invested*, or nearly so, while also never shorting any securities. This means that their portfolio must be carefully rebalanced with each investment decision.

From a mathematical point of view, full investment means that we force the cash investment size

$$w_0 = 0$$

and require all investments to be nonnegative

$$\boldsymbol{w} \geq 0$$

2.1.2. *Partial Short.* A *short* position corresponds in our notation to $w_i < 0$ for some i. Since many securities (such as equities) have no ceiling on their value, there is no ceiling on the losses a short position might theoretically suffer. This has generally meant that shorting is considered a "sophisticated" investment strategy.

As the potential gains to be realized from even a small amount of shorting have come to be commonly known, and the investing public has become (somewhat) better informed, mutual funds have begun to allow their managers to borrow money (also known as using *leverage*).

The most common style, 130/30, will invest up to 160% of notional capital, with up to 130% being long positions and up to 30% being short. The fund is usually constrained to have some metric, such as

gross long minus gross short position, to be 100%.

$$\sum_{i=1}^{N} w_i = 1$$

$$1 \le \sum_{i=1}^{N} |w_i| \le 1.6$$

$$w_0 = 1 - \sum_{i=1}^{N} |w_i|$$

(cash borrowed to fund leverage)

2.2. **Hedge Funds.** From a portfolio management point of view, hedge funds appear very similar to mutual funds. They choose from investable securities, and may operate with some constraints. However, their fee structure and investor cohort ultimately makes them rather different.

With notable exceptions, hedge fund receive about 1% of assets under management (AUM) per year as a *management fee*, plus an *incentive fee* 15% of profits exceeding the highest investment value achieved heretofore (the *high-water mark*). The relatively high management fees should motivate investors to request significant leverage in order to mitigate the effect of the those fees on the Sharpe ratio. Incentive fees, being a proportion of profits, tend to have no effect on the Sharpe ratio.

2.3. **Investors in Funds.**
- Hedge Funds
  - Return-seeking
  - Expect wide asset ranges
  - Loose benchmarks
  - Significant Leverage
- Mutual Funds
  - Safety
  - Want narrow remit
  - Tight benchmarks
  - Negligible Leverage

3. Simple Optimal Portfolios

3.1. **Some Basic Optimization Principles.** Consider an arbitrary function $f(x) : \mathbb{R}^n \to \mathbb{R}$. On any compact set $K \in \mathbb{R}^n$, f will take on a maximum value $f_{\max}(K)$. If f is bounded above and $f \in C^2$ then f has a maximum value $f_{\max}$ which it will achieve at a particular value $x_{\max}$

if f is decreasing outside some compact set K. The same principles clearly apply to minimum values of f.

A value $x_{\max}$ at which f attains a maximum will satisfy the *Kuhn-Tucker Conditions*, particularly the primary *gradient condition* that

$$\nabla f(x_{\max}) = 0$$

This condition is not sufficient, but is necessary. For example, we may have attained a *saddle point* where the *Hessian*, or matrix of second partial derivatives, is neither positive definite nor negative definite. To see that the gradient condition is necessary, simply consider that if the gradient were nonzero in any direction $\boldsymbol{v}$ at $\boldsymbol{x}$, then to first order in δ,

$$f(\boldsymbol{x} + \delta v) = f(\boldsymbol{x}) + \delta|\nabla f| + O(\delta^2)$$

meaning we can find nearby points with more extreme values.

3.2. **Basic Portfolio Modeling.** It is relatively obvious that returns on various market assets are not independent. We begin thinking about the implications of this for a portfolio by making the simplest possible assumption about interdependence of returns, which is that when one return r_1 deviates from its mean, then the other one r_2 tends to do so as well.

Mathematically, we can express this simple demand as follows

$$\mathbb{E}(r_1 - \mu_1 \mid r_2 > \mu_2) \neq \mathbb{E}(r_1 - \mu_1)$$
$$\text{or } Pr(r_1 > \mu_1 \mid r_2 > \mu_2) \neq Pr(r_1 > \mu_1)$$

It makes sense to write the departure from equality as an expectation itself, resulting in our computing

$$\mathbb{E}\big[(r_1 - \mu_1)(r_2 - \mu_2)\big]$$

which is of course just the covariance.

Therefore, we see that the simplest available model must be one in which interdependence of returns is completely characterized by covariance.

3.2.1. *Sums of Random Variables.* If variables

$$X \sim N(\mu_X, \sigma_X)$$

and

$$Y \sim N(\mu_Y, \sigma_Y)$$

are normally distributed with the given means and standard deviations, then their sum $Z = X + Y$ is normally distributed. The sum Z is

distributed as

$$Z \sim N(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$$

It follows that an arbitrary linear combination of normally distributed variables is normally distributed. This property is known as *stability* of the distribution under the additive operator.

The lognormal distribution is not stable., i.e. if X and Y are lognormally distributed then Z is *not* lognormally distributed.

The family of stable distributions is characterized by up to 4 parameters. To keep our model simple for now, we employ the subset of stable distributions characterized by just two parameters. This includes only the uniform interval distributions and the normal distributions. The former are poor representatives of returns, leaving us with normal distributions.

There are two obvious problems with the choice of assuming the r_i are normal.

- Cases where $r_i < 1$ correspond, financially, to losing more than 100% of an investment, which is in general not possible
- We know from both observation and economic theory that dollar returns are more-or-less proportional to asset value. Since returns over any macroscopic period can be perceived as compounded returns over a set of smaller sub periods, a normal model of returns contradicts this observation, suggesting a lognormal model is more appropriate.

3.3. **Optimization Using Lagrange Multipliers.**

3.3.1. *A Too-trivial problem.* Imagine briefly that we had only two available securities, and the full investment constraint

$$w_1 + w_2 = 1$$
$$\boldsymbol{w} \geq 0$$

Let us assume further that we wanted to choose the minimum possible total variance. That is to say, if ρ were the correlation between S_1 and S_2, then we want to minimize

$$\sigma_R^2 = w_1 \sigma_1^2 + w_2 \sigma_2^2 + 2\rho \sigma_1 \sigma_2$$

In this simple case we can substitute for w_2, achieving the problem of minimizing

$$\sigma_R^2 = w_1 \sigma_1^2 + (1 - w_1)\sigma_2^2 + 2\rho \sigma_1 \sigma_2$$

which, being linear in w_1, implies its minimum is either at

$$w_1 = 0$$

or at

$$w_1 = 1$$

3.3.2. *Solving Harder Problems.* Obviously this is a lucky situation, and we will not always have such a simple solution. However, we can use the method of *Lagrange Multipliers* to optimize more difficult constrained problems.

We use the term *feasible* to describe the space in which constraints are satisfied. If the constraints are (linear) equalities then the feasible region is an (affine) subspace S of the domain of f. If they are inequalities then it is a region thereof.

If an extremum occurs in the interior of a subspace, typical solution techniques generally apply, and the Kuhn-Tucker conditions are more or less unaltered. However, on the boundary of a region or in a subspace the situation differs.

Contrained to a subspace, the optimum of f must occur where a change in any direction *within the subspace* would fail to improve the value of f. This is tantamount to requiring that the gradient of f need not necessarily be nonzero, but it had better be zero when projected onto the tangent plane of our subspace. That is to say, it must be perpendicular to the tangent plane of the subspace.

Now, the *contours* or *level sets* of f are necessarily perpendicular to the gradient of f; the gradient of f defines their *normal vectors*. Therefore, if a contour of f is *not* tangent to S then the gradient is not perpendicular to S and the point cannot be an extremum.

Let's say our constraint is expressed with the equation

$$g(\boldsymbol{x}) = 0$$

Then since our normal vectors of the contour and the constraint must be parallel, they are equal to within some factor λ. Therefore we know that for an extreme point,

$$g(\boldsymbol{x}) = 0$$
$$\nabla f(\boldsymbol{x}) = \lambda \nabla g(\boldsymbol{x})$$

We construct a brand-new function in $\Lambda : \mathbb{R}^{n+1} \to \mathbb{R}$ as

$$\Lambda(\boldsymbol{x}) = f(\boldsymbol{x}) - \lambda g(\boldsymbol{x})$$

Requiring that

$$\frac{\partial}{\partial \lambda} \Lambda(\boldsymbol{x}) = 0$$
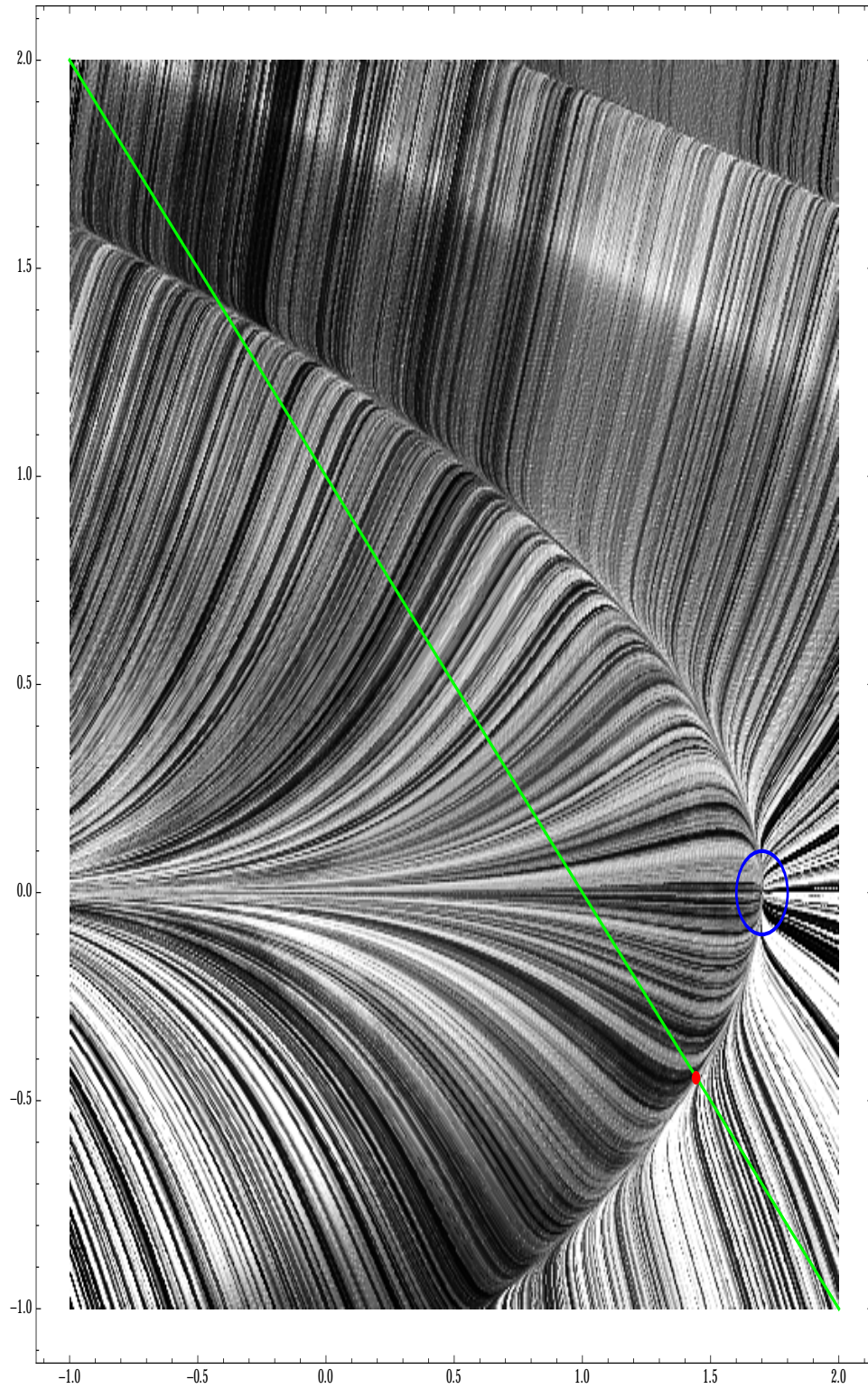
FIGURE 1. Gradient field of $f(x, y) = \frac{6+x}{10} \sin(x + y^2)$

is the same as requiring that

$$g(\boldsymbol{x}) = 0$$

while requiring that (abusing notation somewhat)

$$\frac{\partial}{\partial \boldsymbol{x}} \Lambda(\boldsymbol{x}) = 0$$

is the same as requiring that

$$\nabla f(\boldsymbol{x}) = \lambda \nabla g(\boldsymbol{x})$$

Thus, if we can find λ and $\boldsymbol{x}$ such that

$$\nabla \Lambda(\boldsymbol{x}, \lambda) = 0$$

then we will have found a constrained extremum.

3.3.3. *Concrete Example.* Let's consider the case where

$$f(\boldsymbol{x}) = 2 - x_1^2 - 2x_2^2$$

where our constraint is that $\boldsymbol{x}$ must lie on the unit circle, i.e.

$$g(\boldsymbol{x}) = x_1^2 + x_2^2 - 1 = 0$$

Our Lagrangian is now

$$\Lambda(\boldsymbol{x}, \lambda) = 2 - x_1^2 - 2x_2^2 - \lambda(x_1^2 + x_2^2 - 1)$$

which has a gradient requirement from the partial derivatives

$$\nabla \Lambda(\boldsymbol{x}, \lambda) = \begin{pmatrix} -2x_1 - 2\lambda x_1 \\ -4x_2 - 2\lambda x_2 \\ x_1^2 + x_2^2 - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The first two equations imply $\lambda \in \{-1, -2\}$. The corresponding points are

$$(x_1, x_2, f) = (\pm 1, 0, 1) \quad \text{maxima}$$
$$(x_1, x_2, f) = (0, \pm 1, 1) \quad \text{minima}$$

3.3.4. *Multiple Constraints.* If we have multiple constraints

$$\boldsymbol{g} : \mathbb{R}^K \to \mathbb{R} = (g_k), k = 1, \ldots, K$$

then our gradient must be perpendicular to them all. We end up having to construct λ as a vector $\boldsymbol{\lambda} \in \mathbb{R}^K$ and require that

$$\nabla \Lambda(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$$

where

$$\Lambda(\boldsymbol{x}) = f(\boldsymbol{x}) - \boldsymbol{\lambda}^* \boldsymbol{g}(\boldsymbol{x})$$

One perspective on this is that we have encoded our constraints to find our extremum

$$f^M = \max_x \left(\min_\lambda \left(\Lambda(\boldsymbol{x}, \lambda) \right) \right)$$

where we basically have the power to "move x around" but once we choose a value for it, the worst possible λ will typically send the minimum to $-\infty$. However, we can legitimately change to

$$f^M = \min_\lambda \left(\max_x \left(\Lambda(\boldsymbol{x}, \lambda) \right) \right)$$

which gives us an unconstrained problem and more hope.

In economics, if f is profitability and g represents a resource constraint, then the intuition on the Lagrangian value λ is that it is the *shadow price* of the constraint, representing the amount of profit we might expect for an incremental relaxation of the constraint.

3.4. **Minimal Variance.** For a fully invested minimal variance portfolio, we have the simple target to minimize

$$\sigma_R^2 = \frac{1}{2} \boldsymbol{w}^* \boldsymbol{\Sigma} \boldsymbol{w}$$

subject to our constraint that

$$\boldsymbol{w}^* \boldsymbol{i} = \sum_{i=1}^N w_i = 1$$

We can observe that, if we express the constraint as an equation in an "extra" Lagrange multiplier parameter λ, then the gradient of σ_R^2 must be perpendicular to the level sets of the constraint. Adding level sets of the constraint to our objective function obtains

$$F(\boldsymbol{w}, l) = \frac{1}{2} \boldsymbol{w}^* \boldsymbol{\Sigma} \boldsymbol{w} - l(\boldsymbol{w}^* \boldsymbol{i} - 1)$$

and if we apply the Kuhn-Tucker sufficiency conditions, we obtain

$$\nabla F = \boldsymbol{\Sigma} \boldsymbol{w} - l \boldsymbol{i} = 0$$

This solves to

$$\boldsymbol{w} = l \boldsymbol{\Sigma}^{-1} \boldsymbol{i}$$

Now, knowing that we are on the hypersurface

$$\boldsymbol{w}^* \boldsymbol{i} = 1$$

we can write

$$(l \boldsymbol{\Sigma}^* \boldsymbol{i})^{-1} \boldsymbol{i} = 1$$

or

$$l = \frac{1}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}}$$

Our weight vector is therefore

$$\boldsymbol{w}_{\min} = \left(\frac{1}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{i}$$

$$= \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{i}}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}}$$

Note also we can compute the minimum variance achieved as

$$\frac{1}{2} \boldsymbol{w}_{\min}^* \boldsymbol{\Sigma} \boldsymbol{w}_{\min} = \frac{1}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}}$$

3.5. **Mean/Variance.** If we allow the investor to hold cash, then the portfolio can in principle take on zero risk. Of course, an all-cash portfolio is very difficult to promote to investors (in the long term) as worthy of management fees, so here we consider the problem of optimizing the tradeoff between risk and return.

Aversion to risk might be considered as a proportion of the standard deviation of returns. That is to say, our utility function has a component

$$-\eta \sigma_R$$

It turns out to be more mathematically convenient to assume that our aversion actually is proportional to variance of returns

$$-\lambda \sigma_R^2$$

which is apparently incompatible with aversion to standard deviation. Note that we can series expand around a unknown but existing optimal level σ_M to see that if we define

$$\Delta \sigma = \sigma_R - \sigma_M$$

then

$$-\lambda \sigma_R^2 = -\lambda (\sigma_M + \Delta \sigma)^2$$

$$= -\lambda \sigma_M^2 - 2\lambda \Delta \sigma \sigma_M - \lambda \Delta \sigma^2$$

$$= -\lambda \sigma_M^2 - 2\lambda \Delta \sigma \sigma_M + O(\Delta \sigma^2)$$

$$\approx -\lambda \sigma_M^2 - 2\lambda \Delta \sigma \sigma_M$$

$$= \text{Const} - (2\lambda \sigma_M) \Delta \sigma$$

$$= \text{Const} - (2\lambda \sigma_M) \sigma_R$$

This shows that, up to a constant and errors of polynomial order 2, aversion to variance and aversion to standard deviation are equivalent.

For clarity, we will separate the cash and its weight w_0 from the rest of the weight vector. Let us assume that our aversion to variance is characterized by the parameter λ.

3.5.1. *Optimization With Cash.*

Our constrained optimization problem of maximizing risk-adjusted return becomes to maximize

$$(1) \qquad w_0 r_0 + \boldsymbol{w}^* \boldsymbol{\mu} - \lambda \frac{1}{2} \boldsymbol{w}^* \boldsymbol{\Sigma} \boldsymbol{w}$$

subject to our constraint that

$$(2) \qquad w_0 + \boldsymbol{w}^* \boldsymbol{i} = \sum_{i=0}^{N} w_i = 1$$

$$= w_0 + \sum_{i=1}^{N} w_i$$

Note that if we were to take $\lambda = 0$ then we would be maximizing expected return subject to our overall budget constraint. It's relatively easy to see that in that case the optimal weights set simply puts all the weight on the security with the highest expected return

$$\boldsymbol{w} = \boldsymbol{e}_i \qquad \text{for some } i \leq N$$

In the realistic case, where $\lambda > 0$, we can simply substitute our constraint from Equation (2)

$$w_0 = 1 - \boldsymbol{w}^* \boldsymbol{i}$$

into the objective Equation (1) to achieve the objective function

$$(1 - \boldsymbol{w}^* \boldsymbol{i}) r_0 + \boldsymbol{w}^* \boldsymbol{\mu} - \lambda \frac{1}{2} \boldsymbol{w}^* \boldsymbol{\Sigma} \boldsymbol{w}$$

or

$$r_0 + \boldsymbol{w}^* (\boldsymbol{\mu} - r_0 \boldsymbol{i}) - \lambda \frac{1}{2} \boldsymbol{w}^* \boldsymbol{\Sigma} \boldsymbol{w}$$

Without loss of generality we can leave the constant term r_0 out of the optimization, and create the *excess return* vector of expected returns minus risk-free return level

$$\boldsymbol{\mu}_e = \boldsymbol{\mu} - r_0 \boldsymbol{i}$$

to achieve the unconstrained objective function

$$F(\boldsymbol{w}) = \boldsymbol{w}^* \boldsymbol{\mu}_e - \lambda \frac{1}{2} \boldsymbol{w}^* \boldsymbol{\Sigma} \boldsymbol{w}$$

According to the Kuhn-Tucker conditions, a maximum will occur when $\nabla F(\boldsymbol{w}) = 0$. The gradient is

$$\nabla F(\boldsymbol{w}) = \boldsymbol{\mu}_e - \lambda \boldsymbol{\Sigma} \boldsymbol{w}$$

which therefore requires that

$$\boldsymbol{w}_{\max} = \frac{1}{\lambda} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_e$$

We can compute the cash component as well, since

$$w_0 = 1 - \boldsymbol{w}_{\max}^* \boldsymbol{i}$$

$$= 1 - \frac{1}{\lambda} \boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_e$$

3.5.2. *Fully Invested.* If the portfolio must be fully invested, and cannot use leverage, then we cannot use the cash component as a substitution to solve our equations. We return to Lagrange multipliers to maximize

$$\boldsymbol{w}^* \boldsymbol{\mu} - \lambda \frac{1}{2} \boldsymbol{w}^* \boldsymbol{\Sigma} \boldsymbol{w}$$

which solves to

$$\boldsymbol{w}_{\max} = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{i}}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}} + \frac{1}{\lambda} \frac{\left(\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \left(\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{i}}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}}$$

Now, as $\lambda \to \infty$ we see that the solution is just the minimum variance solution. Also, if all returns are the same, i.e. $\boldsymbol{\mu}$ is a multiple of $\boldsymbol{i}$, then this is the same as the minimum variance solution.

We compute the expected return as

$$\boldsymbol{\mu}^* \boldsymbol{w}_{\max} =$$

$$\frac{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}} + \frac{1}{\lambda} \frac{\left(\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i} \right) \boldsymbol{\mu}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \left(\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right)^2}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}}$$

The expected variance is

$$\boldsymbol{w}_{\max}^* \boldsymbol{\Sigma} \boldsymbol{w}_{\max} =$$

$$\frac{1}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}} + \frac{1}{\lambda^2} \frac{\left(\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i} \right) \boldsymbol{\mu}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \left(\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right)^2}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}}$$

3.5.3. *Long Short Market Neutral.* Let us now consider a fund that must run *dollar neutral*, which is to say that every long investment must be matched by an equivalent short investment. In this case,

maximizing risk-adjusted return is tantamount to solving

$$\boldsymbol{w}^*\boldsymbol{\mu} - \lambda\frac{1}{2}\boldsymbol{w}^*\boldsymbol{\Sigma}\boldsymbol{w}$$

subject to

$$\boldsymbol{w}^*\boldsymbol{i} = 0$$

We solve this to find

$$\boldsymbol{w}_{\max} = \frac{1}{\lambda}\frac{\left(\boldsymbol{i}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{i}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \left(\boldsymbol{i}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{i}}{\boldsymbol{i}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{i}}$$

Let's define

$$\ell = \frac{\boldsymbol{i}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{i}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{i}}$$

then

$$\boldsymbol{w}_{\max} = \frac{1}{\lambda}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \ell\boldsymbol{i}\right)$$

Contrast this with the "standard" case, where we found

$$\boldsymbol{w}_{\max} = \frac{1}{\lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_e$$

$$= \frac{1}{\lambda}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r_0\boldsymbol{i}\right)$$

This shows that the quantity ℓ serves as a control term, wherein higher returns of the standard mean-variance optimal portfolio increase the weights we use and get balanced by cash, and in the dollar neutral portfolio they increase the weights we use and adjust remaining weights in the direction of the Lagrange multiplier, thereby maintaining dollar neutrality.

4. Active Portfolio Management

4.1. **Benchmarks.** It is a sad fact that most money managers do a poor job of choosing investments, to the point that they fail to beat relatively standard benchmarks. It is therefore standard to judge managers against some sort of benchmark. We'll be looking in greater detail at the list of metrics investors apply, but for now we can consider the question, from a portfolio manager's point of view, of controlling for, and minimizing unexpected deviation from, benchmark performance.

4.2. **Active Neutrality.** We can take any quantitative attribute a of a stock, and apply it as a whole to the portfolio by summing it over the entire collection, to compute the *exposure* of our portfolio to the

attribute in question

$$a_{\boldsymbol{w}} = \boldsymbol{w}^* \boldsymbol{a}$$

Let's say this attribute is a measure of desirability. Then as managers we want to hold a lot of it, while removing other contributions to variance. To achieve unit exposure to this attribute while minimizing variance, we can once again optimize, obtaining

$$\boldsymbol{w}_a = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{a}}{\boldsymbol{a}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{a}}$$

Compare with the minimal variance fully-invested portfolio, which has

$$\boldsymbol{w}_{\sigma^2} = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{i}}{\boldsymbol{i}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{i}}$$

The analog is clear – rather than having a constraint

$$\boldsymbol{w}^* \boldsymbol{i} = 1$$

we had required instead that

$$\boldsymbol{w}^* \boldsymbol{a} = 1$$

The individual contributions to risk of the elements of the optimal portfolio can be related to the variance equations. Total variance is

$$\sigma_a^2 = \boldsymbol{w}_a^* \boldsymbol{\Sigma} \boldsymbol{w}_a$$

$$= \frac{1}{\boldsymbol{a}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{a}}$$

which we can multiply out to obtain, for the optimal portfolio,

$$\boldsymbol{a} = \frac{1}{\sigma_a^2} \boldsymbol{\Sigma} \boldsymbol{w}_a$$

If we choose a metric a that happens to be a weighted combination of two other metrics d and f, as in

$$a = \kappa_d d + \kappa_f f$$

then we can compute its characteristic portfolio as a linear combination as well. Define

$$a_d = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{d}}{\boldsymbol{d}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{d}} \boldsymbol{a}$$

$$a_f = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{f}}{\boldsymbol{f}^* \boldsymbol{\Sigma}^{-1} \boldsymbol{f}} \boldsymbol{a}$$

as the exposure of d and f optimal portfolios to the combination a.

Then

$$\sigma_a^2 = \left(\frac{\kappa_d a_d}{\sigma_d^2} + \frac{\kappa_f a_f}{\sigma_f^2} \right)^{-1}$$

and

$$\boldsymbol{w}_a = \kappa_d \frac{\sigma_a^2}{\sigma_d^2} \boldsymbol{w}_d + \kappa_f \frac{\sigma_a^2}{\sigma_f^2} \boldsymbol{w}_f$$

4.2.1. *Benchmark Portfolio.* Let's consider the relationship to a benchmark portfolio. Here we take a given market portfolio M and define "beta" of any other portfolio P as

$$\beta_P = \frac{\text{Cov}(r_P, r_M)}{\text{Var}(r_M)}$$

We write the return of any portfolio as a market return component and an *idiosycratic* component

$$r_P = \beta_P r_M + \theta_P$$

where by the definition of β_P we have that

$$\text{Cov}(\theta_P, r_M) = 0$$

By additivity of variances we can then create the variance equation

$$\sigma_P^2 = \beta_P^2 \sigma_M^2 + \omega_P^2$$

where ω_P^2 is the *idiosyncratic variance*

$$\omega_P^2 = \text{Var}(\theta_P)$$

5. Covariance Sensitivity

Let's study the simple case of mean-variance optimization with risk aversion parameter

$$\lambda = 0.5$$

and covariance matrices computed over trailing 2-month windows on a small set of US equities.

We set the prospective excess returns to be positive for every one of 8 stocks, varying in this case according to a snapshot of price-to-book ratio. Despite the unchanging excess returns, our portfolio weights vary drastically.

This is one of many problems with the Markowitz mean-variance approach, which overall include

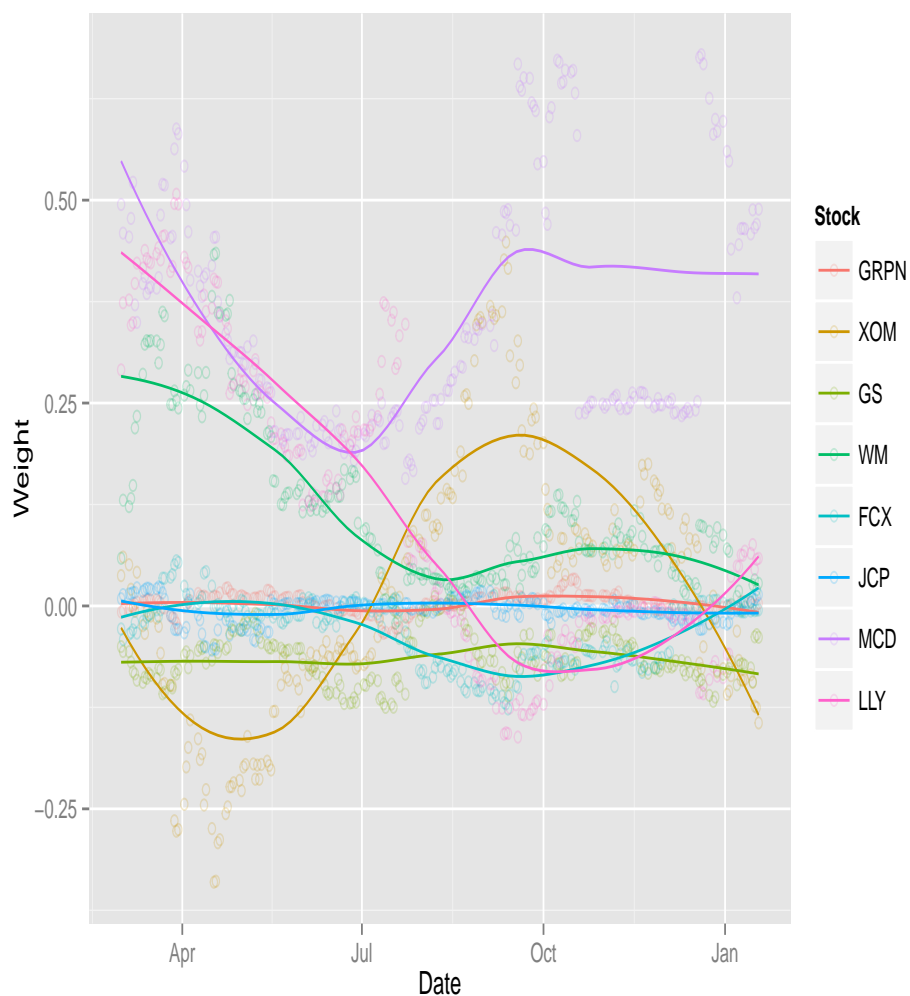- Very sensitive to inputs
- Extreme portfolio weights

FIGURE 2. Unstable Portfolio Weightings

- Huge number of estimated returns
- No idea about confidence intervals
- Single-period
- Variance is two-sided, high returns penalized
- Have to estimate covariance matrix
- Constraint boundaries introduce corners and attract solutions

Ways of dealing with the covariance problem include robust estimation, Ledoit Wolf shrinkage, and enforcing factor structures.

6. Capital Asset Pricing Model

One solution to dealing with the problem of covariance matrices is to drastically simplify the problem. Let's make the (relatively strong) assumption that every return r_i can be decomposed into

$$r_i = r_0 + \beta_i(r_M - r_i) + \epsilon_i$$

where the idiosyncratic ϵ_i are uncorrelated. In this case we can compute

$$\beta_i = \frac{\text{Cov}(r_i, r_M)}{\text{Var}(r_M)} = \rho_{i,M}\frac{\sigma_i}{\sigma_M}$$

The benchmark portfolio with weights $\boldsymbol{b}$ has beta values

$$\boldsymbol{\beta} = \frac{\boldsymbol{\Sigma b}}{\boldsymbol{b}^*\boldsymbol{\Sigma b}}$$

and its variance is

$$\boldsymbol{\Sigma}_C = \boldsymbol{\beta\beta}^*\sigma_M^2 + \text{Diag}(\boldsymbol{\theta} \circ \boldsymbol{\theta})$$

where $\circ$ indicates the *Schur elementwise product*.

6.0.1. *Covariance and Sherman-Morrison.* Taking

$$S = \text{Diag}(\boldsymbol{\theta} \circ \boldsymbol{\theta})$$

$$\boldsymbol{\beta}_s = \boldsymbol{\beta}S^{-1}$$

and

$$\kappa = \sigma_M^2\boldsymbol{\beta}^*(\boldsymbol{\beta}S^{-1})$$

the Sherman-Morrison formula tells us that

$$\boldsymbol{\Sigma}_C^{-1} = S^{-1} - \frac{\sigma_M^2}{1 + \kappa}\boldsymbol{\beta}_s^*\boldsymbol{\beta}_s$$

In the unconstrained risk-averse optimization, we find that the weights must therefore be

$$\boldsymbol{w}_C = \frac{1}{\lambda}\boldsymbol{\Sigma}_C^{-1}\boldsymbol{\mu}_e$$

$$= \frac{1}{\lambda}(S^{-1}\boldsymbol{\mu}_e - \frac{\sigma_M^2}{1 + \kappa}\boldsymbol{\beta}_s^*\boldsymbol{\beta}_s\boldsymbol{\mu}_e)$$

We define a new variable

$$\boldsymbol{w}_S = \frac{1}{\lambda}S^{-1}\boldsymbol{\mu}_e$$

representing the idiosyncratic solution on the idiosyncratic addition risks. Then

$$\frac{1}{\lambda} S^{-1} \boldsymbol{\mu}_e = \boldsymbol{\beta} \circ (\boldsymbol{\mu} S^{-1})$$
$$= \boldsymbol{\beta}^* \boldsymbol{w}_S$$
$$= \beta_0$$

the portfolio beta.

The fully optimal CAPM solution is then

$$\boldsymbol{w}_C = \boldsymbol{w}_S - \frac{\beta_0 \sigma_M^2}{1 + \kappa} \boldsymbol{\beta}_s$$

which for each component works out to

$$\boldsymbol{w}_{i,C} = w_{i,S} - \frac{\beta_0 \sigma_M^2}{1 + \kappa} \frac{\beta_i}{\theta_i^2}$$
$$= w_{i,S} - 1 + \kappa \frac{\mathrm{Cov}(r_i, r_{w_S})}{\theta_i^2}$$

This shows that the optimal weight decreases as the market covariance grows in proportion to idiosyncratic variance.

7. Cointegrated Distributions By Copula

Consider a portfolio of stock, risk-free bonds, CDS, and vanilla equity options held with various counterparties. We want to capture what overall return distribution our portfolio could experience after a period of time T, typically a day, a week, or a month. For example, a sub-portfolio of government bonds worth \$100MM today might be seen to always be worth at least \$90MM even in extreme cases, but if our prime broker (our *counterparty*) defaults then it may be worth zero.

In many cases, particularly for returns of portfolios comprised of basic assets, and with jumps ignored, we can simply estimate a covariance matrix and apply it as for Markowitz theory and its derivatives. Other assets, however, may have highly nonlinear return relationships among themselves. Even a delta-hedged option position is globally rather nonlinear with respect to underlying price, to say nothing of volatility skew.

Let's think about the problem in Monte Carlo terms, both for ease of expression and because the models we generate in this section are intractable by any other numerical approach.

We begin by defining a set of parameters we are willing to consider fundamental in the sense that all security values will be derivable from them. For example, we might choose, as RiskMetrics basically does,

- Overall market return
- Sector return
- Idiosyncratic stock return
- Overall corporate bond spread
- Idiosyncratic bond spread
- Yields on government securities of various tenors
- Overall market implied volatility
- Idiosyncratic implied volatility
- A $\{0, 1\}$ variable for each company, indicating solvency or bankruptcy

From these parameters a given Monte Carlo selection $\vec{\mu}_i$ must allow us to calculate the value of any instrument in our portfolio. For elements of $\vec{\mu}_i$ corresponding to the most of these entries we may be satisfied with a standard gaussian multivariate model, but that fails us with the latter bankruptcy items.

We therefore make use of a more generic *copula* function, allowing the (marginal) distribution of any one element to be of arbitrary shape. For each element j, we convert its actual distribution G to the normal distribution via CDF and inverse CDF functions. That is, we take $x_j \sim G()$ and form a normally distributed $z_j \sim N()$ via

$$z_j := N^{-1}\left(G(x_j)\right)$$

For bankruptcy, we usually work with the Poisson model in which time of bankruptcy $\tau_j \propto e^{-h_j t}$ which also underlies the basic CDS and CDO models used in credit markets.

Since the new variables are all Gaussian, it is now sensible to think of standard multivariate gaussian correlations/covariances. A sample arises from forming a multivariate gaussian draw $\vec{z^{(i)}}$ (with proper covariance structure) and then converting to $\vec{\mu_i}$ via the elementwise inverse of the formula above.

$$x_j := G^{-1}\left(N(z_j^{(i)})\right)$$

For risk computations, we often apply to computational trick of importance sampling to obtain better resolution on risk estimates.

Note that this approach, by using Monte Carlo for each valuation, makes searches for optimal portfolios computationally tedious, but not impossible.

8. Bayesian Approaches to Portfolio Choice: Black-Litterman

Many of the sensitivity and measurement issues around mean-variance portfolio optimization can be addressed by placing asset return forecasts within a Bayesian framework, where we believe our predictions about relative future returns come with varying levels of confidence. Recall the problems with basic Markowitz mean-variance investing, including that it

- Is sensitive to inputs
- Has extreme portfolio weights
- Requires a huge number of estimated returns, one for every security
- Has no idea about confidence intervals
- Works only for a single-period
- Penalizes high returns as much as low ones
- Requires us to estimate a covariance matrix

By including confidence limits, and returns against a baseline, we can begin to address at least the first four issues. The work becomes far less onerous, because we can concentrate our alpha efforts on just a few stocks, and the confidence limits help us place large bets where we want them.

Recall that our Markowitz optimal weights (when cash position are allowed) were given by

$$\boldsymbol{w}_{\max} = \frac{1}{\lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_e$$

where $\boldsymbol{\mu}_e$ represents excess returns. This applies when we are willing (enough) to believe that outcomes are reasonably close to multivariate normal.

If we assume that the "market" itself has properly weighted all allocations, then in a gaussian setting we can use market capitalizations $\boldsymbol{w}_{\mathrm{mkt}} = \boldsymbol{c}$ obtain a corresponding *implied excess return* estimate, analogous to an implied volatility

$$\boldsymbol{\Pi} = \lambda\boldsymbol{\Sigma}\boldsymbol{c}$$

Obtaining a reasonable choice of λ depends, theoretically, on the market's perception of risk which can be inferred by, say, fitting an equilibrium model and computing the corresponding λ

$$\lambda = \frac{r_M - r_{\mathrm{rf}}}{\boldsymbol{c}^*\boldsymbol{\Sigma}\boldsymbol{c}}$$

Or, we can use CAPM, computing

$$\boldsymbol{\Pi} = r_{\mathrm{rf}} + \boldsymbol{\beta}(r_M - r_{\mathrm{rf}})$$

Alternatively, we can notice that the point of this entire exercise is to obtain (Bayesian) modifications to a market benchmark allocation, meaning we require a consistent λ with our own risk appetite.

Our definition of "market" may be the common equity choice of market capitalization, or in the case of other markets a liquidity associated metric. In all cases it should be closely associated with any benchmarks we hope to beat.

We'll then modify the weights according to our beliefs and recalculate an optimal portfolio. We begin by assuming that our prior beliefs are only relevant to a subset of the market, and that they can be expressed in linear combinations of market asset values. We therefore have a *pick matrix* $\boldsymbol{P}$ that selects only values where our prior opinion applies

$$\boldsymbol{P} \cdot \mathbb{E}(r) \sim \Phi(\boldsymbol{v}, \boldsymbol{S})$$

where $\boldsymbol{S}$ represent an uncertainty in our prior beliefs, and typically has entries of zero off the diagonal, or that correspond to cohorts of market outperformance.

We'll choose a parameter τ to denote the relative variance of equilibrium returns versus market expected returns (usually between 0.01 and 0.3, with 0.025 and inverse sample count being common choices, and 1.0 being selected by some related approaches).

If we have complete certainty so that $\boldsymbol{S} = \boldsymbol{0}$ then we are trying to minimize

$$\left(\mathbb{E}(r) - \boldsymbol{\Pi}\right)^* (\tau\boldsymbol{\Sigma}) \left(\mathbb{E}(r) - \boldsymbol{\Pi}\right)$$

constrained to $\boldsymbol{P} \cdot \mathbb{E}(r) = \boldsymbol{v}$ which, using the method of Lagrangians, yields

$$\mathbb{E}(r) = \boldsymbol{\Pi} + \tau\boldsymbol{\Sigma}\boldsymbol{P}^* \left(\boldsymbol{P}\tau\boldsymbol{\Sigma}\boldsymbol{P}^*\right)^{-1} \left(\boldsymbol{v} - \boldsymbol{P}\boldsymbol{\Pi}\right)$$

If we are uncertain we can then compute conditional return distribution by combining gaussian distributions, resulting in the overall distribution conditional on both market expectations $\boldsymbol{\mu}_e$ and prior beliefs $\boldsymbol{v}$, with mean

$$\boldsymbol{\mu}_{\mathrm{BL}} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}^*\boldsymbol{S}^{-1}\boldsymbol{P}\right)^{-1} \left((\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\Pi} + \boldsymbol{P}^*\boldsymbol{S}^{-1}\boldsymbol{P}\boldsymbol{P}^{-1}\boldsymbol{v}\right)$$

and covariance

$$\boldsymbol{\Sigma}_{\mathrm{BL}}^{\mu} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}^*\boldsymbol{S}^{-1}\boldsymbol{P}\right)^{-1}$$

as can be found by running a generalized least squares computation on the market as partitioned into picked and unpicked securities. We can read this as follows:

- $\tau\boldsymbol{\Sigma}$ is a covariance of historical returns. It basically measures confidence in our benchmark, so a tiny τ represents

$$\mathrm{Var}(\mathbb{E}(r)) \ll \mathrm{Var}(r_{\mathrm{hist}})$$

- $\boldsymbol{P}$ encodes our "picks", "prior choices" or "projection"
- $\boldsymbol{S}^{-1}$ is an uncertainty in our estimate quality

Note that as $\boldsymbol{\Sigma} \to \infty$ we obtain $\boldsymbol{\mu}_{\mathrm{BL}} \to \boldsymbol{v}$

More shrinking analysis is at this website.

9. Bayesian Approaches to Portfolio Choice: Copula Opinion Pooling

Let's vaguely define some overall N-dimensional "market" representation $\boldsymbol{M}$, which in the simplest case might be equally weighted average returns over some universe of securities, but in principle could be almost any set of outcomes (e.g. volume weighted credit spreads). We'll take for granted that an acceptable parameterization $F_{\boldsymbol{M}}$ of the probability distribution function $\boldsymbol{M}$ is available either through historical observations or some other mutual agreement.

As in Black-Litterman we will assume that we have some subjective views on a market subset with a pick vector $\boldsymbol{P}$. Each view has a cumulative distribution function

$$\hat{F}_k(v), \qquad k = 1, \ldots, K$$

which will (generically) differ from the market c.d.f. F_k. Given a set of confidences $\boldsymbol{c}$ we define posterior c.d.f. as the weighted averages

$$\tilde{\boldsymbol{F}} = \boldsymbol{c} \circ \hat{\boldsymbol{F}} + (\boldsymbol{1} - \boldsymbol{c}) \circ \boldsymbol{F}$$

Given the market interaction terms for our picked variables, we assume the interactions remain unchanged in the light of the views. That is to say, if two of them were 20% correlated before we applied our views, we assume they remain so. But more generally we can represent the interactions with a *copula* that is not necessarily normal

$$\begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_K \end{bmatrix} \overset{d}{\sim} \begin{bmatrix} F_1(\mu_1) \\ F_2(\mu_2) \\ \vdots \\ F_K(\mu_K) \end{bmatrix}$$

This yields the posterior

$$
\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_K \end{bmatrix} \overset{d}{\sim} \begin{bmatrix} \tilde{F}_1^{-1}(C_1) \\ \tilde{F}_2^{-1}(C_2) \\ \vdots \\ \tilde{F}_K^{-1}(C_K) \end{bmatrix}
$$

Practically, we begin by making a huge market simulation, and computing the marginals $\boldsymbol{F}$ from it. We transform it to a rank matrix, from which we can estimate the copula $\boldsymbol{C}$. We then form marginal posterior c.d.f. from the formulas above by linearly interpolating our univariate simulations into the joint ranking (quantile) functions.