Honey, I Shrunk the Sample Covariance Matrix

Citation: Ledoit, O., & Wolf, M. (2004). Honey, I Shrunk the Sample Covariance Matrix, The Journal of Portfolio Management, 30(4), 110-119

- The sample covariance matrix contains estimation error
  - When N stocks are large, especially relative to the number of historical return observations available, the sample covariance matrix is estimated with a lot of error.
- Covariance matrix shrinkage pulls the most extreme coefficients towards central value to reduce estimation error.
  - Pull down estimated coefficients that are extremely high tend to contain a lot of positive error
  - Low estimated coefficients have a large amount of negative error and need to be pulled upwards.
- Shrinkage reduces tracking error relative to a benchmark index and increases the information ratio.

The Problem

Let w_B = vector of benchmark weights for N stocks
Let x = vector of active weights
$w_p = w_B + x$ = vector of portfolio weights
y = stock return vector
μ = $E(y)$ = vector of expected stock returns
α = $\mu - w_B'\mu$ = vector of expected stock returns
Σ = covariance matrix of stock returns

Expected Returns and Variances

$\mu_B = w_B'\mu$ = expected return on the benchmark
$\sigma_B^2 = w_B'\Sigma w_B$ = variance of the benchmark return
$\mu_P = w_P'\mu$ = expected return of the portfolio
$\sigma_P^2 = w_P'\Sigma w_P$ = variance of the benchmark return
$\mu E = x'\mu$ = expected excess return on the portfolio
$\sigma_E^2 = x'\Sigma x$ = tracking error ~~sth~~ variance

Equalities

$$\left[\mu_P = \mu_B + \mu_E\right] \qquad \left[\sigma_P^2 = \sigma_B^2 + 2w_B'\Sigma x + \sigma_E^2\right]$$

Constraints

Goal: $x'\Sigma x$ => minimize

$w_p' 1 = 1$ => All portfolio weights must add to one
$x'\alpha \geq g$ => The excess return must be greater than or equal to the target "gain", g.
$x' 1 = 0$ => Weights must add to 1
$x \geq -w_B$ => Is a long-only portfolio → so, $w_p \geq 0$
$x \leq c1 - w_B$ => Where c is the upper bound or the maximum position size (i.e. no more than 10% of a given stock is to be allocated).

Shrinkage Principle

- Advantages of S
  - unbiased (i.e. the expected value is equal to the true covariance matrix).
- Disadvantages of S
  - contains a lot of estimation error when there are fewer price datapoints than stocks (i.e. $N \geq T$)
- Shrinkage logic
  - F denotes a highly structured estimator and S denotes a sample covariance matrix
  - Goal is to compromise between a highly structured estimator, F (shrinkage target), and another unstructured estimator, S.
  - Can find a compromise between these two choices through a linear combination. The compromise between these two estimators is determined by the shrinkage coefficient, δ:

$$\delta F + (1-\delta)S \quad , \quad 0 \leq \delta \leq 1$$

- Define the population constant correlation matrix Φ through the population variances and the average population correlation

$$\phi_{ii} = \sigma_{ii} \quad \text{and} \quad \phi_{ij} = \bar{\varrho}\sqrt{\sigma_{ii}\,\sigma_{jj}}$$

- the sample constant correlation matrix F through the sample variances and the average sample correlation
  - if s_{ij}^2 is the sample covariance between two stocks, the shrinkage target, F, is given by:

$$f_{ii} = s_{ii} \quad \text{and} \quad f_{ij} = \bar{r}\sqrt{s_{ii}\,s_{jj}}$$

- Shrinkage constant.

  - 'Optimal' shrinkage constant, δ, between 0 and 1

  ↳ Value that minimizes the distance between the shrinkage estimator and the true covariance matrix (denote δ^*)
  - Estimated optimal shrinkage constant, $\hat{\delta}^*$:

$$\hat{\Sigma}_{\text{Shrink}} = \hat{\delta}^* F + (1 - \delta^*) S$$

- The shrinkage coefficient.

If K was known, we could use K/T as the shrinkage intensity

  - The shrinkage coefficient depends on an estimator $\hat{k}$, where:

$$\hat{k} = \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma}},$$

- Let T denote the number of price datapoints
- Let y_{it} be the returns of the ith security at time t.

- Shrinkage Target

  - The shrinkage target should involve a small number of free parameters (highly structured) and reflect important characteristics of the unknown quantity being estimated

  - Some choices for the shrinkage target could be the single-factor matrix of Sharpe or the constant correlation model: (which is the average of all sample correlations is the estimator of the common correlation).

- Constant correlation model

  - Assumptions:

    Let y_{it}, $1 \leq i \leq N$, $1 \leq t \leq T$, denotes the stock return i during period t

  - Assume that stock returns are independent and identically distributed over time and have finite moments (fourth).

  - Sample the average of the returns of stock i:

$$\bar{y}_i = T^{-1} \Sigma_{t=1}^{T} y_{it}$$

, where Σ denotes the population (true) covariance matrix and S is the sample covariance matrix.

  - Population and sample correlations between stock returns i and j are given by:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \quad \text{and} \quad r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii} s_{jj}}}$$

  - Average population and sample correlations:

$$\bar{\rho} = \frac{2}{(N-1)N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij} \quad \text{and} \quad \bar{r} = \frac{2}{(N-1)N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} r_{ij}$$

- $\hat{\pi}$ estimates the sum of asymptotic variances of the entries of the sample covariance matrix scaled by $\sqrt{T}$:

A consistent estimator for π is

$$\hat{\pi} = \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\pi}_{ij} \quad , \text{ where}$$

$$\hat{\pi}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \left\{ (y_{it} - \bar{y}_{i\cdot})(y_{jt} - \bar{y}_{j\cdot}) - s_{ij} \right\}^2$$

- $\hat{p}$ estimates the sum of asymptotic covariances of the entries of the shrinkage target with the entries of the sample covariance matrix, scaled by $\sqrt{T}$

$$\hat{p} = \sum_{i=1}^{N} \hat{\pi}_{ii} + \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \frac{\bar{r}}{2} \left(\sqrt{\frac{s_{jj}}{s_{ii}}} \, \underline{\hat{\theta}_{ii,ij}} + \sqrt{\frac{s_{ii}}{s_{jj}}} \, \underline{\hat{\theta}_{jj,ij}} \right)$$

where, $\underline{\theta_{ii,ij}} = \frac{1}{T} \sum_{t=1}^{T} \left\{ (y_{it} - \bar{y}_{i\cdot})^2 - s_{ii} \right\} \left\{ (y_{it} - \bar{y}_{i\cdot})(y_{jt} - \bar{y}_{j\cdot}) - s_{ij} \right\}$

and, $\underline{\theta_{jj,ij}} = \frac{1}{T} \sum_{t=1}^{T} \left\{ (y_{jt} - \bar{y}_{j\cdot})^2 - s_{jj} \right\} \left\{ (y_{it} - \bar{y}_{i\cdot})(y_{jt} - \bar{y}_{j\cdot}) - s_{ij} \right\}$

- $\hat{\gamma}$ estimates the misspecification of the (population) shrinkage target

$$\hat{\gamma} = \sum_{i=1}^{N} \sum_{j=1}^{N} (f_{ij} - s_{ij})^2 ,$$

where f_{ij} and s_{ij} are consistent estimators of ϕ_{ij} and σ_{ij}, respectively

- The shrinkage constant is given by $\frac{\hat{k}}{T}$, but this can imply a value greater than 1 or less than 0, we can truncate at 0 or 1 through:

$$\hat{\delta}^* = \max \left\{ 0, \min \left\{ \frac{\hat{k}}{T}, 1 \right\} \right\}$$

- Notes

  - Constant correlation model is not appropriate if the assets came from different asset classes, such as stocks and bonds.

  - Results show that shrinkage beats sample covariance in all scenarios

  - Shrinkage with constant correlation beats single-factor shrinkage for $N \leq 225$

  - Research suggests that adding a constraint on portfolio variance (i.e. $\sigma_p^2 = w_p^T \Sigma w_p$) improves overall efficiency.