Week 11

This week we shall look at a real-world application of RMT in statistics:

Optimisation of large financial portfolios.

Refs:

- Gatheral (2008) "RMT and Covariance estimation".
- Bai, Liu, Wong (2009) "Enhancement of the applicability of Markowitz's portfolio optimisation by utilizing random matrix theory". Mathematical Finance.
- Modern Portfolio theory – Wikipedia.
- Bai, Li, Wong (2013) "The best estimation for high-dim. Markowitz mean variance optimisation".

Modern Portfolio Theory.

Mathematical framework for assembling a portfolio of assets such that the expected return is maximized for a given level of risk.

Risk = Variance.

Dual problems:

- Maximise portfolio expected return s.t. given level of risk
- Minimise risk for a given level of expected return.

- Harry Markowitz 1952.

  $\Longrightarrow$ Nobel Prize 1990.

Problem formulation.

p assets with returns $\mathbb{X} = (x_1, x_2, \cdots, x_p)'$

- Expected returns (ie. Mean) $\mu = \mathbb{E}[\mathbb{X}] = (\mu_1, \cdots, \mu_p)'$
- Covariance matrix $\operatorname{cov} \mathbb{X} = \Sigma = (\sigma_{ij})$.

Investor has capital K ($=1$ WLOG).

Portfolio $\pi = (\pi_1, \pi_2, \cdots, \pi_p)$.

satisfies $\sum_{k=1}^{p} \pi_i = 1$

$V = \sum_{i=1}^{p} \pi_i x_i = \pi' \mathbb{X}$ Portfolio Value.

$R = \mathbb{E} V = \pi' \mu$ Portfolio expected return.

We define risk of portfolio $r = \operatorname{Var}(V) = \pi' \Sigma \pi$

In general, we allow <u>short-selling</u> which means negative weights are allowed for Π.

~~The problem can be posed as a convex optimisation~~ problem (with constraints).

$$\max \ \Pi' \mu$$

$$\text{s.t.} \begin{cases} \Pi' \mathbb{1} = 1. \\ \Pi' \Sigma \Pi \leq \sigma_0^2. \end{cases} \qquad \mathbb{1} = (1, \dots, 1)'$$

Here σ_0^2 is a given level of risk.

The solution Π to problem is called an <u>optimal allocation</u> and expected return $R = \max \Pi' \mu$ is the <u>optimal return</u>.

The problem has a closed-form solution.

<u>Theorem</u> (Markowitz).

(1) If $\dfrac{\mathbb{1}' \Sigma^{-1} \mu \sigma_0}{\sqrt{\mu' \Sigma^{-1} \mu}} < 1$, then

$$R^{(1)} = \sigma_0 \sqrt{\mu' \Sigma^{-1} \mu} \qquad \Pi^{(1)} = \frac{\sigma_0}{\sqrt{\mu' \Sigma^{-1} \mu}} \Sigma^{-1} \mu$$

(2) If $\dfrac{\mathbb{1}'\Sigma^{-1}\mu\sigma_0}{\sqrt{\mu'\Sigma^{-1}\mu}} > 1$ then

$$R^{(2)} = \frac{\mathbb{1}'\Sigma^{-1}\mu}{\mathbb{1}'\Sigma^{-1}\mathbb{1}} + b\left(\mu'\Sigma^{-1}\mu - \frac{(\mathbb{1}'\Sigma^{-1}\mu)^2}{\mathbb{1}'\Sigma^{-1}\mathbb{1}}\right)$$

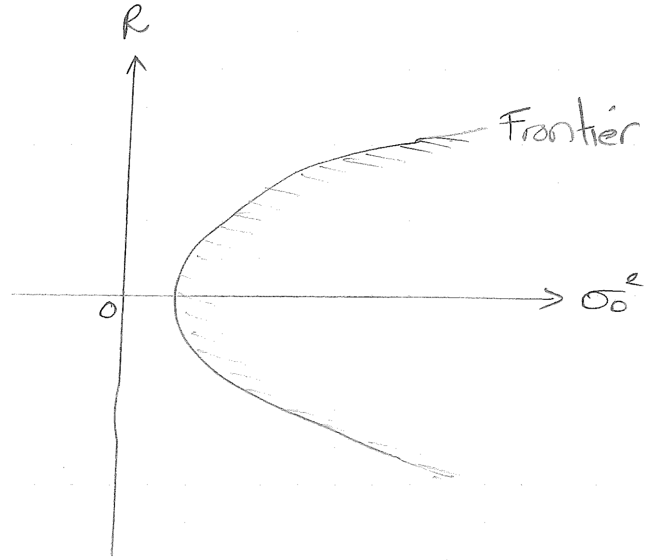
$$\pi^{(2)} = \frac{\Sigma^{-1}\mathbb{1}}{\mathbb{1}'\Sigma^{-1}\mathbb{1}} + b\left(\Sigma^{-1}\mu - \frac{\mathbb{1}'\Sigma^{-1}\mu}{\mathbb{1}'\Sigma^{-1}\mathbb{1}}\Sigma^{-1}\mathbb{1}\right)$$

where $b = \sqrt{\dfrac{\mathbb{1}'\Sigma^{-1}\mathbb{1}\sigma_0^2 - 1}{\mu'\Sigma^{-1}\mu\,\mathbb{1}'\Sigma^{-1}\mathbb{1} - (\mathbb{1}'\Sigma^{-1}\mu)^2}}$ #

<u>Proof:</u> (See Bai et. al. Appendix).

Typically, you solve the problem numerically using an optimisation package. (see Workshop).

The set of optimal portfolios for all possible levels of risk forms the <u>Markowitz mean-variance efficient frontier.</u>



For any given level of risk, there is an optimal return with an optimal portfolio.

These points lie on the frontiér.

The downside is that the Markowitz approach requires knowing:

- Σ (covariance of returns)
- μ (expected returns).

Unfortunately we cannot know their true value so we have to estimate them from data.

Assuming p-returns are observed at n times as

$$\mathbb{X}_i = (x_{i1}, x_{i2}, \cdots, x_{ip}) \qquad i = 1, 2, \cdots, n$$

then (μ, Σ) are estimated by

$$\bar{\mathbb{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{X}_i \qquad \mathbb{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})'$$

Plugging these into the result of the theorem gives us our <u>plug-in estimators</u> of the optimal expected returns and portfolio.

Since the 1950s there has been a massive amount of litterature on Markowitz portfolios.

Many studies have shown that the usefulness of this approach relies heavily on how good your estimate for (μ, Σ) is.

We will look at some recent results that demonstrate (using RMT) why the plug-in portfolio is bad and give a better approach. (Bai et al.)

<u>Over-prediction of returns</u>.

Plugging in $(\bar{\mathbb{X}}, \mathbb{S})$ for (μ, Σ) in the Markowitz theorem gives:

$$\hat{\Pi}_p = \begin{cases} \dfrac{\sigma_0 S^{-1} \bar{\mathbb{X}}}{\sqrt{\bar{\mathbb{X}} S^{-1} \bar{\mathbb{X}}}} & \text{if } \dfrac{\sigma_0 \mathbb{1}' S^{-1} \bar{\mathbb{X}}}{\bar{\mathbb{X}} S^{-1} \bar{\mathbb{X}}} < 1. \\[2ex] \dfrac{S^{-1}\mathbb{1}}{\mathbb{1}' S^{-1} \mathbb{1}} + \hat{b}\left(S^{-1}\bar{\mathbb{X}} - \dfrac{\mathbb{1}' S^{-1} \bar{\mathbb{X}}}{\mathbb{1}' S^{-1} \mathbb{1}} S^{-1}\mathbb{1}\right) & \text{otherwise} \end{cases}$$

where

$$\hat{b} = \sqrt{\frac{\sigma_0^2 \mathbb{1}' S^{-1} \mathbb{1} - 1}{(\bar{\mathbb{X}}' S^{-1} \bar{\mathbb{X}})(\mathbb{1}' S^{-1} \mathbb{1}) - (\mathbb{1}' S^{-1} \bar{\mathbb{X}})^2}}$$

$\hat{\Pi}_p$ is called the <u>plug-in portfolio</u>.

The optimal return R at risk level σ_0^2 is estimated by the plug-in return

$$\hat{R}_p = \hat{\Pi}_p' \overline{\mathbb{X}}$$

This quantity is more useful that $\hat{\Pi}_p'\mu$ since μ is unknown in real-life.

The following theorem (Thm 3.2 Bai et al. 2009) proves that the plug-in approach over-estimates the theoretical return R under a large-dimensional scenario.

Assume observations $\mathbb{X}_1, \mathbb{X}_2, \dots \mathbb{X}_n$ are iid. sample of $\mathbb{X} = \mu + \Sigma^{1/2}\mathbb{y}$ where $\mathbb{y}$ is iid with standard coordinates: $\mathbb{y} = (y_1, \cdots y_p)$, $\mathbb{E} y_i = 0$ and $\mathbb{E} y_i^2 = 1$ and $\mathbb{E} y_i^4 < \infty$.

Theorem: (Bai et al. 2009). Returns $\mathbb{X}_1, \dots \mathbb{X}_n$ satisfy assumption above and $p, n \to \infty$ s.t. $p/n \to y < \infty$. Also the following limit exists:

$$\frac{\mathbb{1}'\Sigma^{-1}\mathbb{1}}{n} \to a_1 \qquad \frac{\mathbb{1}'\Sigma^{-1}\mu}{n} \to a_2$$

$$\frac{\mu'\Sigma^{-1}\mu}{n} \to a_3$$

(cont.)

and satisfy $a_3 a_1 - a_2^2 \geq 0$. Then, almost surely,

$$\lim_{n\to\infty} \frac{\hat{R}_p}{\sqrt{n}} = \begin{cases} \sigma_0\sqrt{\gamma a_3} > \lim\limits_{n\to\infty} \frac{R^{(1)}}{\sqrt{n}} = \sigma_0\sqrt{a_3} & \text{when } a_2 < 0 \\ \sigma_0\sqrt{\gamma(a_3 - a_2^2/a_1)} > \lim\limits_{n\to\infty} \frac{R^{(2)}}{\sqrt{n}} = \sigma_0\sqrt{(a_3 - a_2^2)/a_1} & \text{when } a_2 > 0. \end{cases}$$

Where $R^{(1)}$ and $R^{(2)}$ are theoretical returns given in Markowitz theorem and $\gamma = 1/(1-y) > 1$. //

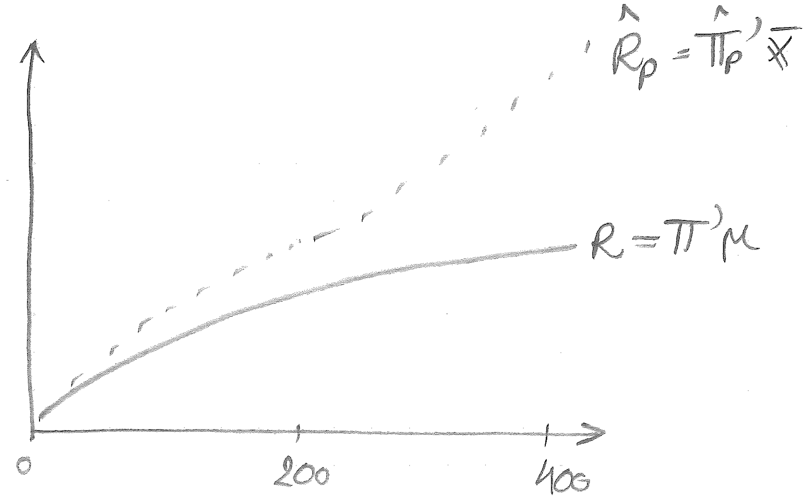
<u>Proof:</u> (See Bai et al. 2009).

Notice that in both cases

$$\lim_{n\to\infty} \frac{\hat{R}_p}{R^{(j)}} = \sqrt{\gamma} > 1 \qquad j = 1, 2.$$

Which gives precise meaning to "over-prediction".

<u>Fig 4.1 in Bai</u>



In Bai et al. 2009, they propose a "bootstrap correction method". (See that paper).

Spectrum-corrected estimator

We will now look at the approach from the 2013 technical report by Bai, Li and Wong.

We need an estimator of Σ.

Assume that when p is large, the eigenvalues of Σ satisfy the spectral decomposition

$$\Sigma = U \Lambda U'$$

$$\Lambda = \mathrm{diag}(\underbrace{\lambda_1, \lambda_1, \dots, \lambda_1}_{p_1}, \underbrace{\lambda_2, \dots, \lambda_2}_{p_2}, \dots \underbrace{\lambda_L, \dots, \lambda_L}_{p_L})$$

with L distinct eigenvalues of respective multiplicity (p_j).

We partition the eigenvector matrix $U = (U_{p_1}, U_{p_2}, \dots, U_{p_L})$ so that

$$\boldsymbol{\Sigma} = \sum_{j=1}^{L} \lambda_j U_{p_j} U_{p_j}' .$$

We are going to assume that as $p \to \infty$,

$$\frac{p_j}{p} \to w_j > 0 \qquad j = 1, \cdots, L.$$

In other words, we have a population spectral distribution (PSD) of the form

$$H = \lim_{p\to\infty} H_p = \sum_{j=1}^{L} w_j \delta_{\lambda_j}$$

There are a few RMT techniques to estimate the PSD from observations $\mathbb{X}_1, \mathbb{X}_2, \cdots \mathbb{X}_n$.

For example, Bai, Chen, Yao (2010).
Li, Chen, Qin, Bai, Yao (2013)

<u>Algorithm:</u> (see p. 15, Bai, Li, Wong 2013 & Li, Chen, Qin, Bai Yao 2013)

(1) Set $B = \frac{1}{n} \mathbb{X}\mathbb{X}'$ $\quad \mathbb{X} = (\mathbb{X}_1, \cdots \mathbb{X}_n)$

(2) Compute eigenvalues of B: $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_p$

(3) Choose $\{u_1, \cdots, u_m\} \subset (-\infty, \lambda_1) \cup (\lambda_p, \infty)$ $\quad m \ge p$

for each u_i compute

$$\underline{s}_n(u_i) = -\frac{1-y}{u_i} + \frac{1}{n}\sum_{j=1}^{p} \frac{1}{\lambda_j - u_i}$$

and plug pairs $(u_i, \underline{s}_n(u_i))$ in MP equation

given by
$$u = -\frac{1}{\underline{s}} + y\int \frac{t}{1+t\underline{s}}\,dH(t)$$

---

to get the approximate equations (m of them)
$$u_j \approx -\frac{1}{\underline{s}_n(u_j)} + y\int \frac{t}{1+t\underline{s}_n(u_j)}\,dH(t,\theta) =: \hat{u}_j(\underline{s}_{nj},\theta)$$

Where $H = H(\theta)$ is the limit of H with unknown parameter vector $\theta \in \mathbb{R}^q$.

(4) Find the least squares solution of θ,
$$\hat{\theta}_n = \arg\min_{\theta} \sum_{j=1}^{m} \left(u_j - \hat{u}_j(\underline{s}_{nj},\theta)\right)^2$$

$\hat{\theta}_n$ is the least-squares estimator.

Eg. $h(t;\theta) = \dfrac{2(1-\theta)^2}{(t-a)} \mathbb{1}(t \geq \theta)$ $\quad a := 2\theta - 1.$

$\theta \to 1^-$ then $h \to F_y$ (MP density $H = \delta_1$).

See Li, Chen, Qin, Bai, Yao 2013 for other examples.

We assume $H = \sum_{j=1}^{L} \omega_j \delta_{\lambda_j}$

~~and estimate~~ $\cancel{\hat{\Theta} = \{(\hat{\omega}_j, \hat{\lambda}_j);\ j = 1, \dots, L\}}$ ~~using~~
the algorithm.

Let $S_n = VDV'$ be the spectral decomposition of the sample covariance matrix where V is the orthogonal matrix formed by the eigenvectors.

Let $\hat{\Theta} = \{(\hat{\omega}_j, \hat{\lambda}_j) : j = 1, \dots, L\}$ be the estimators of the PSD parameters.

The <u>spectrum corrected estimator</u> of Σ is

$$\hat{\Sigma}_S = V \hat{\Lambda} V'$$

$$\hat{\Lambda} = \mathrm{diag}(\underbrace{\hat{\lambda}_1, \dots \hat{\lambda}_1}_{\hat{p}_1}, \underbrace{\hat{\lambda}_2, \dots, \hat{\lambda}_2}_{\hat{p}_2}, \dots, \underbrace{\hat{\lambda}_L, \dots, \hat{\lambda}_L}_{\hat{p}_L})$$

$$\hat{p}_j := p \cdot \hat{\omega}_j \qquad j = 1, \dots, L.$$

Notice that this estimator is made of V from SVD of S_n and estimator $\hat{\Lambda}$ of spectrum of Σ.

Since Δ has a finite number of parameters and $\hat{\Delta}$ is a consistent estimator of Δ, the asymptotic properties of

$$\hat{\Sigma}_S = V\hat{\Delta}V'$$

is largely identical to

$$B_p = V\Delta V'$$

so we can study the behavior of B_p instead of $\hat{\Sigma}$

Spectrum corrected estimates for the optimal return and π

Plug in $(\bar{X}, \hat{\Sigma}_S)$ for (μ, Σ) in Markowitz theorem. to get

$$\hat{\pi}_S = \begin{cases} \dfrac{\sigma_0 \hat{\Sigma}_S^{-1}\bar{X}}{\sqrt{\bar{X}'\hat{\Sigma}_S^{-1}\bar{X}}} & \text{if } \dfrac{\sigma_0 \mathbb{1}'\hat{\Sigma}_S^{-1}\bar{X}}{\sqrt{\bar{X}'\hat{\Sigma}_S^{-1}\bar{X}}} < 1. \\[2ex] \dfrac{\hat{\Sigma}_S^{-1}\mathbb{1}}{\mathbb{1}'\hat{\Sigma}_S^{-1}\mathbb{1}} + \hat{b}_S\left(\hat{\Sigma}_S^{-1}\bar{X} - \dfrac{\mathbb{1}'\hat{\Sigma}_S^{-1}\bar{X}}{\mathbb{1}'\hat{\Sigma}_S^{-1}\mathbb{1}}\hat{\Sigma}_S^{-1}\mathbb{1}\right) & \text{if } \dfrac{\sigma_0 \mathbb{1}'\hat{\Sigma}_S^{-1}\bar{X}}{\sqrt{\bar{X}'\hat{\Sigma}_S^{-1}\bar{X}}} > 1 \end{cases}$$

where $\hat{b}_S = \sqrt{\dfrac{\mathbb{1}'\hat{\Sigma}_S^{-1}\mathbb{1}\sigma_0^2 - 1}{\bar{X}'\hat{\Sigma}_S^{-1}\bar{X} \cdot \mathbb{1}'\hat{\Sigma}_S^{-1}\mathbb{1} - (\mathbb{1}'\hat{\Sigma}_S^{-1}\bar{X})^2}}$

The portfolio estimator $\hat{\pi}_S$ is the <u>spectrum-corrected portfolio</u>

The corresponding spectrum-corrected return is

$$\hat{R}_S = \hat{\Pi}_S' \bar{X}$$

and spectrum corrected risk is $\hat{r}_S = \hat{\Pi}_S \hat{\Sigma}_S^{-1} \mu$

Notice that the behaviour of scalar products

$$\mathbb{1}' \hat{\Sigma}_S^{-1} \mathbb{1} \quad , \quad \mathbb{1}' \hat{\Sigma}_S^{-1} \mu \quad , \quad \mu' \hat{\Sigma}_S^{-1} \mu$$

will determine the asymptotic properties of the spectrum-corrected estimators of optimal return and portfolio.

Let $\underline{a} = (a_p)_{p \geq 1}$ and $\underline{b} = (b_p)_{p \geq 1}$ be two sequences of unit vectors where for each p, a_p and b_p are p-dim. vectors.

Assume Σ has a finite PSD with eigenvalues (λ_j) and eigenvector matrices $\{ U_{p_j}, U_{p_j}' \}$.

The sequences $\underline{a}$ and $\underline{b}$ are called Σ-stable if

$$\lim_{p \to \infty} a_p' U_{p_j} U_{p_j}' b_p = d_j \qquad j = 1, \ldots, L$$

The limits $\underline{d} = \{ d_j \}$ are called Σ-characteristics of the pair $(\underline{a}, \underline{b})$

A Σ-stable pair $(\underline{a}, \underline{b})$ is such that the inner products between their projections onto the L eigenspaces of Σ tend to a limit.

This implies.

$$\lim_{p\to\infty} \underline{a}_p' \Sigma^{-1} \underline{b}_p = \sum_{k=1}^{L} \frac{d_k}{\lambda_k}$$

Theorem (Bai, Li, Wong [Thm 4.3])

With $B_p = V \Lambda V'$, $\underline{a}, \underline{b}$ Σ-stable. with Σ-char $\underline{d}$

$p, n \to \infty$ $\quad p/n \to y \in (0, \infty)$.

(1) almost surely

$$a_p' B_p^{-1} b_p \longrightarrow \xi_H(\underline{d}) := \sum_{k=1}^{L} \frac{d_k}{\lambda_k} \sum \frac{\lambda_k (u_j - \lambda_j)}{\lambda_j (u_j - \lambda_k)}$$

for $j = 1, \dots, L$ and u_j is a solution of

$$1 + y \int \frac{t}{u-t} \, dH(t) = 0.$$

satisfying $\lambda_1 > u_1 > \lambda_2 > \dots > \lambda_L > u_L > 0$.

(2) If the projections $U_{p_j} U_{p_j}' a_j$ and $U_{p_j} U_{p_j}' b_p$ on the L eigenspaces of Σ only have finite nonzero entries, it holds almost surely.

$$a_p' B_p^{-1} \Sigma B_p^{-1} b_p \longrightarrow \rho_H(\underline{d}) = \sum_{k=1}^{L} \frac{d_k}{\lambda_k} \left(\sum_{j=1}^{L} \frac{\lambda_k (u_j - \lambda_j)}{\lambda_j (u_j - \lambda_k)} \right)^2$$

Corollary For $\hat{\Sigma}_S$ it holds that

$$a_p' \hat{\Sigma}_S^{-1} b_p \longrightarrow \xi_H(\underline{d})$$

$$a_p' \hat{\Sigma}_S^{-1} \Sigma \hat{\Sigma}_S^{-1} b_p \longrightarrow \rho_H(\underline{d})$$

(see Thm 4.5)

For a vector v let $v_0 = v/\|v\|$ be projection onto unit sphere.

Theorem: Assume

$$(\mathbb{1}_0, \mathbb{1}_0), \ (\mathbb{1}_0, \mu_0) \text{ and } (\mu_0, \mu_0)$$

are Σ-stable with Σ-char $\underline{d}_1$, $\underline{d}_2$, and $\underline{d}_3$.

Set $\xi_j = \xi_H(\underline{d}_j)$ for $1 \leq j \leq 3$.

Assume $p \to \infty$, $\|\mu\| = \xi_1(1 + o(1))$ for $\xi_1 > 0$. Then as $p, n \to \infty$ $p/n \to y \in (0,1)$

(1) $\mathbb{1}_0' \hat{\Sigma}_S^{-1} \mathbb{1}_0 \to \xi_1$ $\quad \mathbb{1}_0' \hat{\Sigma}_S^{-1} \mu_0 \to \xi_2$ $\quad \mu_0' \hat{\Sigma}_S^{-1} \mu_0 \to \xi_3$

$\mathbb{1}_0' \hat{\Sigma}_S^{-1} \bar{X}_0 \to \xi_2$, $\quad \bar{X}_0' \hat{\Sigma}_S^{-1} \bar{X}_0 \to \xi_3$

(2) Almost surely,

$$\hat{R}_S \simeq \begin{cases} \sigma_0 \xi_1 \rho \sqrt{\xi_3} & \text{if } \sigma_0 \sqrt{\rho}\, \xi_2 / \xi_3 < 1 \\ \xi_1 \dfrac{\xi_2}{\xi_1} + \xi_1 \left(\rho \sigma_0^2 - 1/\xi_1 \right)^{1/2} \left(\xi_3 - \xi_2^2 / \xi_1 \right)^{1/2} & \text{if } \sigma_0 \sqrt{\rho}\, \xi_2 / \xi_3 > 1 \end{cases}$$

//