Chapter 11

Portfolio Management

11.1 Introduction and Problem Settings

11.1.1 Returns

The chapter of portfolio management is concerned with the management of a basket of capital assets. For instance, the price of a government bond or stock at time t can be denoted p_t. This is an example of a capital asset. The pnl of holding one unit of the asset at $[t-1, t)$ is $\pi_t = p_t - p_{t-1}$. The price process is a stochastic process. The price at any time t is then some initial condition plus a sum of profits, $p_t = p_l + \sum_{j=l+1}^{t} \pi_j$. If π_t is IID with $\mathbb{E}\pi_t = \mu, \text{Var}(\pi) = \sigma^2$, we call this drift and variance respectively and the price process is random walk. Then the conditional mean $\mathbb{E}[P_t|P_0] = P_0 + t\mu$ and conditional variance is $\text{Var}(P_t|P_0) = \sigma^2 t$. Continuous-time treatment of an asset price process sees asset prices as (generalized) geometric Brownian motion (see Definition 113). We do not deal with continuous-time treatment here.

We can define returns in many ways, and here we shall define some. The *gross return*, also known as cumulative returns is simply $G_t(k) = \frac{P_t}{P_{t-k}}$, the ratio of prices over two periods. We assume positive price processes, $G_t(k) \geq 0$. The k period gross return is simply the product of 1-period gross returns (or more generally, product of gross returns on non-overlapping periods). This is expressed $G_t(k) = \Pi_{i=1}^{k} G_{t-k+i}(1)$. The *net return* is $R_t(k) = G_t(k) - 1 = \frac{P_t - P_{t-k}}{P_{t-k}}$. The *log return* is $r_t(k) = \log G_t(k) = \log(P_t) - \log(P_{t-k})$. It is often true that we do not distinguish between net returns and log returns in literature. The true relationship is

$$r_t(k) = \log(1 + R_t(k)) \approx R_t(k), \tag{789}$$

where the approximation is given by $\log(1+x) \approx x$ when x is small (here log is the natural log, 'ln') . Their distinction is of little concern in practice when working on fine granularities of data. However, when working with returns over longer day periods, this approximation does not necessarily hold. The mathematics is simple but the implications are non-trivial. Log-returns are often favored - it is quite well behaved over time. The k-period log return is additive over the 1-period log returns (again, any non-overlapping period holds) such that $r_t(k) = \sum_{i=1}^{k} r_{t-k+i}(1)$. Most often, we will not distinguish between daily net returns and daily log return data and use them interchangeably.

We are interested in the economics of multiple assets. These assets put together form a portfolio. Let there be p assets, and w_i be the weight assigned to asset $i \in [p]$. A general portfolio allocation can be denoted $\|w\|_1 = \sum_i^p |w_i| = 1$. If we are working with strategies, or positive return bearing assets, it makes sense to constrain our weight to positive values, such that $\forall i, w_i \geq 0$. Let it be the index for asset i at time t. Then

$$P_t = \left(1 + \sum_{i=1}^{p} w_i R_{it}\right) P_{t-1}, \tag{790}$$

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \sum_{i=1}^{p} w_i R_{it}, \tag{791}$$

$$r_t = \log(1 + \sum_{i=1}^{p} w_i R_{it}). \tag{792}$$

The naive but highly useful assumption is that $r_t \sim \Phi(\mu, \sigma^2)$. Then $\log(P_t) - \log(P_{t-k}) \sim \Phi(k\mu, k\sigma^2)$. This is the discrete form statement for asset price processes driven by a Brownian motion. For continuous time-treatments, see Definition (141). Log-normality is also obtained there. In particular, the asset price is modelled

$$P_t = P_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma w_t\right\}, \qquad w_t \sim \Phi(0, t). \tag{793}$$

Often the portfolio returns are taken in relation to a benchmark. We call this excess returns. For some asset price benchmark p'_t, the excess returns are $r_t - r'_t$. Often the benchmark is set to the risk-free rate. Cash rates or short-term (3M) Treasury rates are often used in literature when computing Sharpe ratios. There are many objections. Firstly, benchmarks are arbitrary and choices are themselves questionable. Secondly, most people live in the nominal world. When we talk about our portfolio returns at a barbeque party, we say: 'we made $x\%$ a year'. If you say: 'I made $y\%$ net of YoY core CPI inflation', then feel free to benchmark the risk-free rate. Just don't come to my party.

When growth rates (such as interest rates) are continuously compounded, we get nice approximations. If for some small time period $\frac{1}{n}$ we compound at rate $\frac{r}{n}$, then after t periods we have

$$P_t = P_0 \left(1 + \frac{r}{n}\right)^{nt} \overset{n\to\infty}{\to} P_0 \exp(rt). \tag{794}$$

These have applications in discounting future valuations of wealth, such as cash flows. When the rates are allowed to be stochastic, then we get discount *processes*. See Definition 2023 for continuous time treatments. It would not very important at this juncture.

11.1.2 Risk

Almost everything we want to know about portfolio management is with regards to estimating the true distribution, or nature of r_t. The first moment and second central moment (see Definition 40) for r_t is the expected return and variance respectively, namely

$$\mu = \mathbb{E}r_t, \qquad \sigma^2 = \mu_2 - \mu_1^2 = \mathbb{E}[(r_t - \mu)^2]. \tag{795}$$

$\sigma = \sqrt{\sigma^2}$ is known as standard deviation (statistics terminology), volatility (finance terminology) or portfolio risk (trader terminology). There are many definitions of portfolio risk, but volatility is by far the most commonly used. Some commonly used distributions in the study of returns are normal distributions (Section 5.17.4), t-distribution (Section 5.17.7) and log-normal distributions (when r_t is exponentiated, see Section 5.17.6).

Result 13. *If $\hat{\theta} - \theta \sim \Phi(0, \frac{1}{n}\xi^2)$, then for any $f(\theta)$ with $|f'(\theta)| < \infty$ we have*

$$f(\hat{\theta}) - f(\theta_0) \sim \Phi\left(0, \frac{1}{n}\left(f'(\theta_0)\right)^2 \xi^2\right). \tag{796}$$

Lemma 14. *With IID sample $Y_i, i \in [n]$, for $\hat{\mu} = \bar{Y}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_i^n (Y_i - \bar{Y})^2$, then*

$$\begin{aligned} \sqrt{n}(\hat{\mu} - \mu) &\sim \Phi(0, \sigma^2), && (797) \\ \sqrt{n}(\hat{\sigma}^2 - \sigma^2) &\sim \Phi(0, 2\sigma^4), && (798) \\ \sqrt{n}(\hat{\sigma} - \sigma) &\sim \Phi(0, \frac{1}{2}\sigma^2) && (799) \end{aligned}$$

Proof. The first two results follow immediately from Definition 84. For the third statement, consider the Taylor expansion of $(\hat{\sigma}^2)^{\frac{1}{2}}$ at σ^2, expressed

$$\begin{aligned} \hat{\sigma} &= (\hat{\sigma}^2)^{\frac{1}{2}} \approx (\sigma^2)^{\frac{1}{2}} + \frac{1}{2(\sigma^2)^{\frac{1}{2}}}(\hat{\sigma}^2 - \sigma^2) && (800) \\ &= \sigma + \frac{1}{2\sigma}(\hat{\sigma}^2 - \sigma^2). && (801) \end{aligned}$$

Since $(\hat{\sigma}^2 - \sigma^2) \sim \Phi(0, \frac{2\sigma^4}{n})$, apply Result 13 with $f(x) = x, \xi = \sqrt{2}\sigma^2$ to get

$$\hat{\sigma} \sim \Phi\left(\sigma, \frac{1}{4\sigma^2}\left(\frac{1}{n}2\sigma^4\right)\right) \tag{802}$$

and we are done. □

11.1.2.1 VaR, Conditional VaR

Definition 133 (Value at Risk)**.** *Recall the definition of random variable quantiles (see Definition 26). Let the q quantile be denoted $Q_q(x)$. For continuous c.d.f we can also write $Q_q(x) = F^{-1}(q)$. When the random variable under concern is r_t, then $-Q_q(X)$ is q VaR. Write*

$$VaR_q(X) = -Q_q(X) = -\max\{x : F(x) \leq q\}. \tag{803}$$

This is the minimum loss incurred in the worst q fraction of return samples.

Lemma 15 (Properties of VaR)**.** *Let X, Y be random variables, and $\lambda > 0, c \in \mathbb{R}$. The following properties are satisfied by VaR risk measures.*

1. $VaR_q(X + c) = VaR_q(X) - c$.
2. $X \leq Y \implies VaR_q(X) \geq VaR_q(Y)$. *VaR is consistent with first-order stochastic dominance.*
3. $VaR_q(\lambda X) = \lambda VaR_q(x)$.

In particular, $VaR_q(a + bX) = bVaR_q(X) - a$ when $b > 0$.

Proof. Proof for part 1. Write $VaR_q(X) = -Q_q(X)$ is such that

$$\begin{aligned} \mathbb{P}(X < Q_q(X)) &= q && (804) \\ \mathbb{P}(X + c < Q_q(X) + c) &= q && (805) \end{aligned}$$

but $-Q_q(X) - c$ is $VaR_q(X + c)$. For part 2, see

$$\begin{aligned} \mathbb{P}(X < Q_q(X)) = q && (806) \\ \mathbb{P}(X + (Y - X) < Q_q(X) + (Y - X)) = q && (807) \\ \mathbb{P}(Y < Q_q(X) + (Y - X)) = q && (808) \end{aligned}$$

but $-Q_q(X) + X - Y$ is $VaR_q(Y)$. $X - Y < 0$ and the result follows. □

Exercise 100. *For $X \sim \Phi(\mu, \sigma^2)$ we know that for z_q, the q-th quantile of $\Phi(0,1)$, we have*

$$\mathbb{P}\left(\frac{X-\mu}{\sigma} < z_q\right) = q \tag{809}$$
$$\mathbb{P}(X < \mu + z_q\sigma) = q \tag{810}$$

s.t. $Q_q(x) = \mu + z_q\sigma, VaR_q(x) = -\mu - z_q\sigma$. The same principle applies if we assume $X \sim t_v$-distribution and so on. If we fit a distribution to r_t, we can the find the VaR at arbitrary q using the estimated parameters. If we have sufficient data, we can estimate based on empirical quantiles to high resolution without making parametric assumptions. (see Definition 26).

Definition 134 (Conditional VaR, Expected Shortfall)**.** *Recall the definition of value-at-risk in Definition 133. This measured the minimum loss incurred. The expected shortfall measures the average loss incurred in the worst q fraction of return samples instead, expressed*

$$ES_q(X) = \frac{1}{q}\int_0^q VaR_\alpha(X)d\alpha. \tag{811}$$

Why conditional VaR? See

$$\begin{aligned} ES_q(X) &= -\mathbb{E}[X|X < -VaR_q(X)] & (812)\\ &= -\frac{1}{q}\int_{-\infty}^{-VaR_q(X)} xf(x)dx & (813)\\ &= -\frac{1}{q}\int_{-\infty}^{-VaR_q(X)} xdF(x) & (814)\\ &= -\frac{1}{q}\int_0^q F^{-1}(\alpha)d\alpha \qquad \alpha = F(x) & (815)\\ &= \frac{1}{q}\int_0^q VaR_\alpha(X)d\alpha. & (816)\end{aligned}$$

Lemma 16 (Properties of Expected Shortfall)**.** *It is easy to see using the results from Lemma 15 that*

1. $ES_q(X+c) = ES_q(X) - c$, $c \in \mathbb{R}$
2. $X \le Y \implies ES_q(X) \ge ES_q(Y)$.
3. $ES_q(\lambda X) = \lambda ES_q(X)$. $\lambda > 0$.

11.1.2.2 Risk-Adjusted Returns

The portfolio manager is interested in returns, particularly in relation to risk. He should be interested primarily in the Sharpe ratio, which is computed as the ratio of expected returns to volatility of returns. Most often, the annualized Sharpe ratio is presented. Suppose daily returns r_t are independent. Then let $R = \sum_i^n r_i$. Then $\text{Var}(R) = \sum_i^n \text{Var}(r_i) = n\sigma_r^2$ and $\sigma_R = \sqrt{n}\sigma_r$. Suppose we have daily return data, and let number of trading days in a year be 253, then the annualized Sharpe is computed

$$\begin{aligned} sharpe &= \frac{253 * \mu_r}{\sqrt{253}\cdot\sigma_r} & (817)\\ &= \sqrt{253}\cdot\frac{\mu_r}{\sigma_r}. & (818)\end{aligned}$$

The reason why the Sharpe ratio is of primary concern can be outlined. It gives us the return measured in units of volatility. Volatility can be acquired by leverage. A high Sharpe portfolio with low risk can be levered up to a desired level of volatility. For given level of volatility, the returns possible are commensurate with the Sharpe ratio. As mentioned, some choose to use benchmarked returns. Others criticize the Sharpe ratio all together. For the naysayers of Sharpe, I leave with you a quote (Paleologo [5]):

> `committees have been formed; replacements have been suggested, including the`
> `beautifully named 'ulcer index'; recommendations have been ignored.`
> `Titans of finance come and are quickly forgotten.`
> `Volatility and Sharpe will stay in the foreseeable future.`

We show important results in the confidence of sample statistics on Sharpe ratio observable on return data from market trading.

Corollary 20 (Sharpe Ratio)**.** *We may estimate the Sharpe ratio by $\hat{SR} = \frac{\hat{\mu}}{\hat{\sigma}}$ with distribution*

$$\hat{SR} - SR \sim \Phi\left(0, \frac{1}{n}\left(1 + \frac{1}{2}SR^2\right)\right). \tag{819}$$

The 95% *confidence interval for the Sharpe Ratio is*

$$\hat{SR} = \hat{SR} \pm 1.96\sqrt{\frac{1}{n}\left(1 + \frac{1}{2}\hat{SR}^2\right)} \tag{820}$$

This gives us the range of values for which a hypothesis testing for zero Sharpe would return statistically (in)significant p-values.

Proof. By bivariate Taylor expansion arguments, we argued in Lemma 41 that the variance of ratio $R = \frac{Y}{X}$ is given by

$$\text{Var}(R) \approx \frac{1}{\mu_x^2}(r^2\sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2 - 2r\sigma_{\bar{x}\bar{y}}) \tag{821}$$

See from Result 14 that we have $\hat{\sigma} \sim \Phi(\sigma, \frac{\sigma^2}{2n})$, and $\hat{\mu} \sim \Phi(\mu, \frac{\sigma^2}{n})$. Ignoring the finite population correction factor, we can write our bootstrap approximate of the Sharpe estimator variance

$$\text{Var}(SR) \approx \frac{1}{\hat{\sigma}^2}\left(SR^2\frac{\hat{\sigma}^2}{2n} + \frac{\hat{\sigma}^2}{n}\right) \qquad \hat{\sigma} \perp \hat{\mu} \text{ , Definition 84} \tag{822}$$

$$= \frac{1}{n}\left(\frac{1}{2}SR^2 + 1\right) \tag{823}$$

and we are done. ☐

11.2 A Basket of Assets

Assets are held in aggregate, and this is called a portfolio. The portfolio wealth process is observed with respect to a basket of assets evolving over time. Market instruments are intricately related, and so are the random variables representing the observables of interest. These random variables shall then be studied with multivariate methods. In the general form, for $X = (x_i)_{i\in[p]}$ random variables, the random variables have c.d.f $F(x_1, x_2 \cdots, x_p) = \mathbb{P}(X_1 < x_1 <, \cdots X_p < x_p)$. If there exists joint density, then their relations are given by

$$\mathbb{P}(X_1 \in (a_1, b_1), \cdots X_p \in (a_p, b_p)) = \int_{a_1}^{b_1} \cdots \int_{a_p}^{b_p} f(x_1, \cdots x_p) dx_p \cdots x_1. \tag{824}$$

The mathematics of joint densities are discussed in Section 59. Of particular prevalence is the multivariate Gaussian distributions (see Definition 85). Recall that they are characterized completely by their expectations and covariance matrix, which we denote here (μ, Σ). Refer to Definitions of random variable expectations, covariance and correlation in Definitions 31, 44 and 44 respectively. In particular, for p-column matrix $Y = (Y_i)_{i\in[p]}^T$ we have

$$\begin{aligned} \mathbb{E}(Y) &= (\mathbb{E}Y_i)_{i\in[p]}^T = (\mu_i)_{i\in[p]}^T, && (825) \\ Cov(Y) = \Sigma &= \mathbb{E}\left\{[Y - \mathbb{E}Y][Y - \mathbb{E}Y]^T\right\}. && (826) \end{aligned}$$

The covariance and correlation matrices are said to be semi-positive definite. More generally (obtained by their definitions and using the transpose $(AB)^T = B^T A^T$ property):

$$\mathbb{E}AY = A\mathbb{E}Y, \qquad Cov(AY) = ACov(Y)A^T. \tag{827}$$

More generally, $Cov(AY, BY) = ACov(Y)B^T$. We can estimate the population covariance matrix with the observed data. For observed Y, take

$$S = \frac{1}{n}\sum_{l=1}^{n}(Y_l - \bar{Y})(Y_l - \bar{Y})^T. \tag{828}$$

To reduce the variance of this estimator, we may apply shrinkage methods while trading off for bias (see Definition 111). For shrinkage value λ, the shrinkage (ridge) estimator can be written

$$S' = (1 - \lambda)S + \lambda\mathbb{1}, \tag{829}$$

where the λ choice is made by cross-validation. Other alternatives, such as the Ledoit-Wolf shrinkage can be employed.

11.2.1 Computations for the Portfolio

For p assets with returns $R = (R_i)_{i\in[p]}$, denote their expected returns and covariance as $r_i = \mathbb{E}R_i, Cov(R) = \Sigma$ respectively. The portfolio is a linear combination of assets - let the weighting be denoted by (column) vector w, s.t the portfolio R_p has relation $R_p = w^T R$. In the general case, $\|w\|_1 = 1$, but this is usually $\sum w_i = 1$ (components are non-negative) in the strategy-allocation problem (since we would not be trading a strategy with negative expected returns). By the relation given by Equation 827, the portfolio's expected returns, variance, volatility given by

$$r_p = \mathbb{E}R_p = w^T\mathbb{E}R, \qquad \sigma_p^2 = w^T\Sigma w, \qquad \sigma_p = \sqrt{w^T\Sigma w}. \tag{830}$$

It follows that the portfolio Sharpe is given

$$SR_p = \frac{w_T r_p}{\sqrt{w^T\Sigma w}}. \tag{831}$$

Our objectives would be to maximise r_p, minimize σ_p, or their ratio SR_p. The efficient frontier curve is the curve along non Pareto dominated solutions for w w.r.t to the first two objectives. In particular, it is the set of solutions for w corresponding to (i) maximum r_p for some target σ_p, or to (ii) minimum σ_p for some target r_p. To trade on this frontier would be an efficient approach - see otherwise that the trader's SR_p would be lower.

11.2.2 Elegant Mathematics, Poor Economics

We present the classical mathematical models of optimization. These are known as 'Modern Portfolio Theory'. There is nothing quite modern about these models, and no serious amount of capital are managed by the mathematics presented herein. However, the advanced models of today would be motivated by similar goals of optimization. The theory presented herein are the bedrock for evolutions in portfolio theory.

We present the arguments here, assuming constraint $\sum w_i = 1$, with no short-constraints. This is to simplify the mathematics under the settings of 'soft' constraints.

Exercise 101 (Global Minimum Variance)**.** *Assume p assets, with returns $R = (R_i)^T_{i\in[p]}$ and covariance $\Sigma = (\sigma_{ij})_{i\in[p],j\in[p]}$. The global minimum variance portfolio is the portfolio $R_p = w^T R$ that solves*

$$\begin{aligned} \min_w Cov(R_p) &= w^T \Sigma w, && (832) \\ s.t. \sum_p w_i &= 1. && (833) \end{aligned}$$

Proof.

$$Cov(R_p) = w^T\Sigma w = \sum_l^p w_l^2 \sigma_l^2 + 2\sum_{i<j}^p w_i w_j \sigma_{ij}. \tag{834}$$

Lagrangian for the problem is

$$L(w,\lambda) = \sum_l^p w_l^2\sigma_l^2 + 2\sum_{i<j}^p w_i w_j \sigma_{ij} + \lambda(\sum_l^p w_i - 1) = 0, \tag{835}$$

with Lagrangian equations

$$\frac{\delta L(w,\lambda)}{\delta w_i} = 2\sum_j w_j \sigma_{ij} + \lambda = 0, \tag{836}$$

$$\frac{\delta L(w,\lambda)}{\lambda} = \sum_i w_i - 1 = 0. \tag{837}$$

This can be written by block matrix

$$\begin{bmatrix} 2\Sigma & \mathbb{1} \\ \mathbb{1}^T & 0 \end{bmatrix} \begin{bmatrix} w \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{838}$$

Solving for the set of linear equations $2\Sigma w = -\lambda\mathbb{1}$, $\mathbb{1}^T w = 1$, we get

$$\begin{aligned} w &= -\frac{1}{2}\Sigma^{-1}\lambda\mathbb{1}, && (839) \\ \mathbb{1}^T w &= -\frac{1}{2}\mathbb{1}^T\Sigma^{-1}\lambda\mathbb{1} = 1. && (840) \end{aligned}$$

It follows that $\lambda = -2\frac{1}{\mathbb{1}^T\Sigma^{-1}\mathbb{1}}$, with solution

$$w = \frac{\Sigma^{-1}\mathbb{1}}{\mathbb{1}^T\Sigma^{-1}\mathbb{1}}. \tag{841}$$

□

Exercise 102 (Markowitz Portfolios)**.** *The efficient portfolio is set up*

$$\max r_p = w^T r \tag{842}$$
$$s.t. \quad w^T \Sigma w = \sigma^2_{target} \quad \& \quad \sum_i w_i = 1. \tag{843}$$

or

$$\min Cov(r_p) = w^T \Sigma w \tag{844}$$
$$s.t. \quad w^T r = r_{target} \quad \& \quad \sum_i w_i = 1. \tag{845}$$

Proof. Solving the second optimization problem form, we get Lagrangian

$$L(w, \lambda_1, \lambda_2) = w^T \Sigma w + \lambda_1 (w^T r - r_{target}) + \lambda_2 (w^T \mathbb{1} - 1), \tag{846}$$

with Lagrangian equations

$$\frac{\delta L}{\delta w} = 2\Sigma w + \lambda_1 r + \lambda_2 \mathbb{1} = 0 \tag{847}$$
$$\frac{\delta L}{\delta \lambda_1} = w^T r - r_{target} = 0 \tag{848}$$
$$\frac{\delta L}{\delta \lambda_2} = w^T \mathbb{1} - 1 = 0. \tag{849}$$

Solving, we get (verify this)

$$w = \frac{c - b r_{target}}{ac - b^2} \Sigma^{-1} \mathbb{1} + \frac{a r_{target} - b}{ac - b^2} \Sigma^{-1} r, \tag{850}$$

where $a = \mathbb{1}^T \Sigma^{-1} \mathbb{1}, b = \mathbb{1}^T \Sigma^{-1} r, c = r^T \Sigma^{-1} r.$ □

Exercise 103 (Maximum Sharpe/Tangency Portfolio)**.** *Solve for*

$$\max_w \frac{w^T r}{(w^T \Sigma w)^{\frac{1}{2}}} \tag{851}$$

subject to $w^T \mathbb{1} = 1$.

Proof. Optimization equation has Lagrangian

$$L(w, \lambda) = \frac{w^T r}{(w^T \Sigma w)^{\frac{1}{2}}} + \lambda (w^T \mathbb{1} - 1). \tag{852}$$

The first order conditions are

$$\frac{\delta L}{\delta w} = r (w^T \Sigma w)^{-\frac{1}{2}} - (w^T r)(w^T \Sigma w)^{-\frac{3}{2}} \Sigma w + \lambda \mathbb{1} = 0, \tag{853}$$
$$\frac{\delta L}{\delta \lambda} = w^T \mathbb{1} - 1 = 0. \tag{854}$$

Solving, (verify this) we get

$$w = \frac{\Sigma^{-1} r}{\mathbb{1}^T \Sigma^{-1} r}. \tag{855}$$

□