Applications to Finance

We will look at 2 application areas:

1) Portfolio construction

2) Risk managment

Machine Learning Asset Allocation

Mean-Variance Optimization

Consider a market with n assets

Let r_i denote the simple returns of the i^{th} asset

↳ a random variable

We form a portfolio of these n assets using weights w_i

↳ The return of the portfolio is $R = w_1 r_1 + w_2 r_2 + \dots + w_n r_n$

If vector of returns r has mean μ $(E[r_i] = \mu_i)$
and covariance C $(\mathrm{cov}(r_i, r_j) = c_{ij})$
then the portfolio has expected return $E[R] = \mu^T w = w_1\mu_1 + w_2\mu_2 + \cdots + w_n\mu_n$
and variance $\mathrm{Var}(R) = w^T C w = \sum_i \sum_j w_i w_j c_{ij}$

The Markowitz model: maximize returns + minimize risk

↳ $\min \frac{1}{2} w^T C w - \lambda \mu^T w$ subject to $\sum_{i=1}^{n} w_i = 1$

where $\lambda \geq 0$ encodes the level of risk aversion (parameterizes the optimal portfolio)

Sometimes add other constraints, ex: long-only: $w_i \geq 0$ for all i

This solution can be computed very efficiently (FE 630)

However, it relies on accurate estimates of μ + C

↳ even small errors in μ + C can lead to very different "optimal" portfolios

→ the computed portfolio is suboptimal

Markowitz's Curse

1) In general, we need $\frac{1}{2}n(n+1)$ i.i.d. observations to estimate the covariance matrix of n assets (and even then we get the sample $\hat{C} \neq C$)

<u>ex</u>: $n = 50$ assets will require at least 5 years of daily data

$n = 100$ assets will require at least 20 years of daily data

2) In computing the optimal portfolio, we need to invert the covariance matrix C^{-1}
When assets are more correlated, this inverse can introduce numerical errors

$\Rightarrow$ The more correlated the assets, the greater the need for diversification
& the more likely we will recieve unstable solutions

Viewing Markowitz as Machine Learning

Recall: $\min \frac{1}{2} w^T C w - \lambda \mu^T w$ subject to $\sum_{i=1}^{n} w_i = 1$

This optimal portfolio is a function of 3 main inputs: $w = f(\mu, C, \lambda)$

(Assume we know λ, drop it from notation)

In reality, we find: $\hat{w} = f(\hat{\mu}, \hat{C})$

↪ $\hat{w}$ typically performs very poorly on test data (out-of-sample)

↳ Due to imperfect estimation + large amount of data required

↳ Typically outperformed by the naive investment $w_i = \frac{1}{n}$ for every asset

Using rolling window to estimate $\hat{\mu}$ + $\hat{C}$ then:

$\hat{w}_t = f_T(r_{t-T}, \ldots, r_t)$ is a function of realized historical returns

This is a complex regression problem

↳ Training outputs are historical optimal portfolios

Still very difficult to implement reliably!

Classification Approach

Take notion from naive portfolio ($w_i = \frac{1}{n}$ or 60-40 split of equity-bond) to choose to only invest in the entire market (SPY) or "risk-free" (IEF)

Consider additional predictors as well:

ex:

- VIX index
- value of gold
- sector ETFs
- moving average of prior data
- trade volume
- ...

Want to predict the "direction" of SPY returns

↳ $y_t = \begin{cases} 1 & \text{if } r_t^{SPY} \geq \tau \\ 0 & \text{if } r_t^{SPY} < \tau \end{cases}$ for threshold returns τ

Approach 1

Directly run classification methods on this data set

If predicted "1" then invest in SPY for next period

If predicted "0" then invest in IEF for next period

Approach 2

Use the "probability" of being in class 0/1

Idea: Invest in SPY for next period when predicted probability of class 1 exceeds threshold K

↳ allows you to select how confident you want to be in the market

$\underline{ex}$: $\tau = -1\%$, $K = 90\%$

↳ avoid SPY if greater than 10% chance of 1% (or more) drop in the market

(can be generalized to multiple assets, but more complex)

Quantile Regression + Risk Analysis

3 Typical Measures of Risk

1) Variance
2) Value-at-Risk (VaR) ←
3) Expected Shortfall (ES)

Value-at-Risk is a measure of the risk of loss for investments

Estimates how much an investment might lose with a given probability (in a set period of time)

If X is the profit & loss distribution of a portfolio, then VaR at level $\alpha \in (0,1)$ is the $(1-\alpha)$-quantile of $-X$

↳ $$\mathrm{VaR}_\alpha(X) = -\inf\left(x \in \mathbb{R} \mid F_X(x) > \alpha\right)$$

↳ The 95%-VaR provides the losses in the worst 5% of situations

Estimating VaR

Historical Perspective

Idea: Use changing or rolling window to get the empirical quantile

↳ find the value so that α fraction of profits/losses are better

Collect data on market variables for T days in the past

This provides T alternative scenarios for what can happen between today + tomorrow

Use these returns to compute the dollar change in the value of the portfolio between today + tomorrow

↳ Value under the t^{th} scenario $= V_T \, {}^{V_t}/_{V_{t-1}}$

Quantile Regression

Rather than purely use historical data to find the quantile of V_t/V_{t-1} we can try to <u>regress</u> this value

<u>Goal</u>: Find a loss function so that we estimate the quantile of the signal

<u>Recall</u>: The α-quantile of X is: $q_\alpha(X) = \inf(x \in \mathbb{R} \mid F_X(x) > \alpha)$
and $VaR_\alpha(X) = -q_\alpha(X)$

Assume X has a continuous distribution

[<u>Intuition for MSE</u>: $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ minimizes $\frac{1}{N}\sum_{i=1}^{N}(y_i - t)^2$

↪ to test: take the derivative of the loss function: $-\frac{2}{N}\sum_{i=1}^{N}(y_i - t) = 0$

↪ This is why the MSE leads to a regression of the mean]

Let's start the intuition to find a potential loss function

↳ Define the loss function $\rho_\alpha(x) = x\left(\alpha - \mathbb{1}_{\{x<0\}}\right)$

↳ $\rho_\alpha(X)$ minimizes $\mathbb{E}\left[\rho_\alpha(X-t)\right] = (\alpha-1)\int_{-\infty}^{t}(x-t)\,dF_X(x) + \alpha\int_{t}^{\infty}(x-t)\,dF_X(x)$

↳ test by taking the derivative : $-\alpha + F_X(x) = 0$

For <u>α-quantile regression</u> we use the ρ_α loss function:

$$\hat{\beta}_\alpha = \underset{\beta}{\arg\min}\ \frac{1}{N}\sum_{i=1}^{N}\rho_\alpha\left(y_i - f(x_i,\beta)\right)$$

$$= \underset{\beta}{\arg\min}\ \frac{1}{N}\left[(\alpha-1)\sum_{y_i<f(x_i,\beta)}\left(y_i - f(x_i,\beta)\right) + \alpha\sum_{y_i>f(x_i,\beta)}\left(y_i - f(x_i,\beta)\right)\right]$$

for parametric model $f(\cdot,\beta)$

↳ Commonly used for linear models : $f(x,\beta) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$

but works for more complex models as well (<u>ex</u>. neural nets) ex: $\alpha = \frac{1}{2}$ then $\rho_\alpha(x) = \frac{1}{2}|x|$

↳ median regression minimizes the mean absolute error

Evaluating Model Performance

Throughout this course, we have generally compared loss functions directly to determine the best fit

It is better to statistically test this comparison instead

↳ Only use a more complex model if it has a statistically significant improvement in performance

Elicitability (terminology primarily used in risk measurement)

There exists some scoring function that can be used for comparative tests on models

↳ S is such that $\mathbb{E}[S(\rho(Y), Y)] < \mathbb{E}[S(t, Y)]$ for $t \neq \rho(Y)$

risk evaluation of Y

Let f + g be 2 models of risk

f is better than g w.r.t. S if $\frac{1}{N}\sum_{i=1}^{N} S(f(x_i), y_i) < \frac{1}{N}\sum_{i=1}^{N} S(g(x_i), y_i)$

Note: the relation between scoring functions + loss functions

Value-at-Risk is elicitable

(Expected Shortfall is not, though it is if jointly evaluated with VaR)

One-Sided Test

the "standard" or baseline model

$H_0: \mathbb{E}[S(f(X), Y)] = \mathbb{E}[S(g(X), Y)]$

$H_a: \mathbb{E}[S(f(X), Y)] < \mathbb{E}[S(g(X), Y)]$

Define $\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i$ for $d_i = S(f(x_i), y_i) - S(g(x_i), y_i)$

Under the null hypothesis: $\bar{d}$ is asymptotically normal with mean 0 and standard deviation $\hat{\sigma}_n / \sqrt{n}$