Algorithmic Trading & Quantitative Strategies

Lecture 3 (2/27/2024)

Giuseppe Paleologo (gardener)

Today's Session

- I will upload the slides (only) tonight after class
- There will be a short multiple-answer quiz at the end of the lecture.
- Please send answers in the form
  - 1-a
  - 2-c
  - ...

To gp2642@nyu.edu. You will have 5 minutes

Topics

- Recap from Last Class
  - Factor Models
- Today:
  - How do you evaluate a model? (Important!)
  - Fundamental (characteristic) model estimation

Factor Models: $\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t$

1. **Interpretations:** Graphical Model, Overlaps, Scalar Products

2. **Operations:** Project (simplify), Rotate (change the perspective), Push Out (add variables)

3. **Applications:** Risk Analysis, Portfolio Management, Alpha Research, Performance Attribution

Basic Portfolio Construction

- Mean-Variance is ubiquitous. Why?
- Exponential utility, arbitrary return distribution
  - No wealth effect, constant risk aversion; very unintuitive
- Quadratic utility
  - It is not monotonically increasing!
  - If one only has mean and variance, the only utility consistent with expected utility theory must be quadratic
- Locally quadratic utility
  - Seems a very reasonable approximation

Essential Formulation

- N.B.: All models use excess returns, i.e., returns minus the risk-free rate
  - What risk-free? Many choices: OBFT, SOFR, Three-Month Treasurys? Usually the latter
  - Why? Because it's more economical to think this way. Every position is self-funding: borrow at risk-free, buy security
  - Does not matter for dollar-neutral portfolio
  - In the real world funding rate is not uniform: borrow costs of shorts vary a lot

Essential Formulations (cont.)

This
$$\max \boldsymbol{\alpha}'\mathbf{w}$$
$$\text{s.t. } \mathbf{w}'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w} \leq \sigma^2$$

Or this
$$\max_{\mathbf{w}} \frac{\boldsymbol{\alpha}'\mathbf{w}}{\sqrt{\mathbf{w}'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w}}}$$

Equivalent, because

$$\max_{\mathbf{w}} \frac{\boldsymbol{\alpha}'\mathbf{w}}{\sqrt{\mathbf{w}'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w}}}$$
$$\text{s.t. } \sqrt{\mathbf{w}'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w}} \leq \sigma$$
$$\text{equivalent to } \max_{\mathbf{w}} \frac{\boldsymbol{\alpha}'\mathbf{w}}{\sigma}$$
$$\text{s.t. } \mathbf{w}'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w} \leq \sigma^2$$
$$\text{equivalent to } \max_{\mathbf{w}} \boldsymbol{\alpha}'\mathbf{w}$$
$$\text{s.t. } \mathbf{w}'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w} \leq \sigma^2$$

Additional Formulations

These are all the same

$$\max \boldsymbol{\alpha}'\mathbf{w} - \lambda \mathbf{w}'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w}$$

$$\min \mathbf{w}'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w}$$
$$\text{s.t. } \boldsymbol{\alpha}'\mathbf{w} \geq \mu$$

Sharpe Ratio

$$\mathrm{SR}^\star = \sqrt{\boldsymbol{\alpha}'\boldsymbol{\Omega}_{\mathbf{r}}^{-1}\boldsymbol{\alpha}}$$

Precision matrix

$$\mathbf{w}^\star = \frac{1}{2\lambda}\boldsymbol{\Omega}_{\mathbf{r}}^{-1}\boldsymbol{\alpha}$$

Partial correlations

$$w_i \propto [\Omega_{\mathbf{r}}^{-1}]_{i,i}\left(\alpha_i - \sum_{j \neq i} \rho_{i,j}[\Omega_{\mathbf{r}}^{-1}]_{j,j}\alpha_j\right)$$

Trading in Factor Space

What are the portfolios that best track a factor return? Think about it:

- Portfolio should have 0 exposure to all factors other than i
- And minimum idio vol

$$
\begin{aligned}
&\min \mathbf{w}'\mathbf{\Omega}_\epsilon \mathbf{w} \\
&\text{s.t. } \mathbf{B}'\mathbf{w} = \mathbf{e}_i
\end{aligned}
$$

Investing in Factors

Formulas: $$\mathbf{P} := (\mathbf{w}_1 | \ldots | \mathbf{w}_m) = \mathbf{B}(\mathbf{B}'\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}^{-1}\mathbf{B})^{-1}(\mathbf{e}_1 | \ldots | \mathbf{e}_m) = \mathbf{B}(\mathbf{B}'\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}^{-1}\mathbf{B})^{-1}$$

Expected returns: $(\boldsymbol{\alpha} \perp' + \boldsymbol{\lambda}'\mathbf{B}')\mathbf{w}_i = \lambda_i$

And then you can build a portfolio using MVO on synthetic portfolios. This is what a more sophisticated strategy would do:

$$\frac{\boldsymbol{\Omega}_{\mathbf{f}}^{-1}\boldsymbol{\lambda}}{\sqrt{\boldsymbol{\lambda}'\boldsymbol{\Omega}_{\mathbf{f}}^{-1}\boldsymbol{\lambda}}}$$

Trading in Idio Space

Factor-neutralized portfolios:

$$\begin{aligned} &\max \boldsymbol{\alpha}_{\perp}' \mathbf{w} \\ &\text{s.t. } \mathbf{B}'\mathbf{w} = 0 \\ &\qquad \mathbf{w}'\boldsymbol{\Omega}_{\epsilon}\mathbf{w} \leq \sigma^2 \end{aligned}$$

And recall the lower bound:

$$\mathrm{SR}(\mathbf{w}) \geq \frac{\|\boldsymbol{\alpha}_{\perp}\|}{\|\boldsymbol{\Omega}_{\epsilon}\|_{\mathrm{op}}} \geq \sqrt{n}\frac{\mu}{\|\boldsymbol{\Omega}_{\epsilon}\|_{\mathrm{op}}}$$

Live Overview of Factors

(Live)

Investment: Metrics

- Sharpe is the first metric being analyzed because high SR is necessary, not sufficient, for strategy deployment
- Capacity is extremely important to portfolio managers
- What you want is to **maximize PnL subject to a SR constraint**
- A bound on SR also bounds the tails:
  - With normal returns: $P(\xi < -L\sigma)=F^{-1}(-L-SR)$
  - With finite variance: $P(\xi < -L\sigma) \leq \dfrac{1}{1+(L+\mathrm{SR})^2}$ (Cantelli's lemma)
  - Reality: in between.
- In practice, VaR and CVaR are not as used by the buy-side as much as the sell-side (for regulatory purposes)
  - Unstable (esp CVaR), so they can't be used as denominator and for risk budgeting purposes
  - Effectively, they rely on higher moments, and we can't estimate these moments
  - Their theoretical benefits are only theoretical

Fundamental Law of Active Investing

- Makes rigorous the relationship between diversification and performance
- Start with MVO portfolio: $\mathbf{w}^\star = \frac{\sigma}{\sqrt{\boldsymbol{\alpha}'\boldsymbol{\Omega}_\mathbf{r}^{-1}\boldsymbol{\alpha}}}\boldsymbol{\Omega}_\mathbf{r}^{-1}\boldsymbol{\alpha}$
- Expected return transformations:

$$
\begin{aligned}
E(\mathbf{r}'\mathbf{w}^\star) &= \frac{\sigma}{\sqrt{\boldsymbol{\alpha}'\boldsymbol{\Omega}_\mathbf{r}^{-1}\boldsymbol{\alpha}}} E(\mathbf{r}'\boldsymbol{\Omega}_\mathbf{r}^{-1}\boldsymbol{\alpha}) \\
&= \sigma \frac{E(\mathbf{r}'\boldsymbol{\Omega}_\mathbf{r}^{-1}\boldsymbol{\alpha})}{\sqrt{\boldsymbol{\alpha}'\boldsymbol{\Omega}_\mathbf{r}^{-1}\boldsymbol{\alpha}}\sqrt{E(\mathbf{r}'\boldsymbol{\Omega}_\mathbf{r}^{-1}\mathbf{r})}} \sqrt{\frac{E(\mathbf{r}'\boldsymbol{\Omega}_\mathbf{r}^{-1}\mathbf{r})}{n}} \sqrt{n}
\end{aligned}
$$

FLAM

Define the **Information Coefficient:**

$$\mathrm{IC} := \frac{E(\mathbf{r}'\mathbf{\Omega}_{\mathbf{r}}^{-1}\boldsymbol{\alpha})}{\sqrt{\boldsymbol{\alpha}'\mathbf{\Omega}_{\mathbf{r}}^{-1}\boldsymbol{\alpha}}\sqrt{E(\mathbf{r}'\mathbf{\Omega}_{\mathbf{r}}^{-1}\mathbf{r})}}$$

$$\Rightarrow$$

$$\begin{aligned}\tilde{\mathbf{r}} &= \mathbf{\Omega}_{\mathbf{r}}^{-1/2}\mathbf{r}\\ \tilde{\boldsymbol{\alpha}} &= \mathbf{\Omega}_{\mathbf{r}}^{-1/2}\boldsymbol{\alpha}\end{aligned}$$

$$\mathrm{IC}(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{r}}) := \frac{E(\tilde{\mathbf{r}}'\tilde{\boldsymbol{\alpha}})}{\sqrt{\tilde{\boldsymbol{\alpha}}'\tilde{\boldsymbol{\alpha}}}\sqrt{E(\tilde{\mathbf{r}}'\tilde{\mathbf{r}})}}$$

$$\mathrm{SR} = \mathrm{IC}(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{r}}, \mathbf{\Omega}_{\mathbf{r}})\sqrt{n}$$

Sharpe Ratio Statistics

Important. You will use this all the time

$$\hat{\mu} := \frac{1}{T}\sum_{t=1}^{T} r_t \qquad \hat{\sigma} := \sqrt{\frac{1}{T}^{-1}\sum_{t=1}^{T}(r_t - \hat{\mu})^2}$$

$$\widehat{\mathrm{SR}} := \frac{\hat{\mu}}{\hat{\sigma}}$$

In practice, only two formulas:

iid returns, asymptotic formula

$$\mathrm{SE}(\widehat{\mathrm{SR}}) = \sqrt{\frac{1 + \mathrm{SR}^2/2}{T}}$$

Autocorrelated returns AR(1): $r_{t+1} = \rho r_t + \varepsilon_t$

$$\widehat{\mathrm{SR}} := \frac{\hat{\mu}}{\hat{\sigma}}\sqrt{\frac{1-\rho}{1+\rho}} \simeq \frac{\hat{\mu}}{\hat{\sigma}}\sqrt{\frac{1-\rho}{1+\rho}} \simeq \frac{\hat{\mu}}{\hat{\sigma}}(1-\rho)$$

The relationship between Regression and IR

Being a correlation, the IC is also naturally related to the predictive strength of our alphas, as measured by a cross-sectional regression. As an important step in exploring alphas, we perform a cross-sectional weighted-least squares regression of residual returns against alpha. We estimate a coefficient x that solves the following minimization problem:

$$\min_b \sum_i \frac{(r_i - \alpha_i b)^2}{\sigma^2_{\epsilon,i}} = \min_b \|\tilde{\mathbf{r}} - \tilde{\boldsymbol{\alpha}} b\|^2 \tag{4.56}$$

The solution is given by $b^\star = (\tilde{\mathbf{r}}'\tilde{\boldsymbol{\alpha}})/(\tilde{\boldsymbol{\alpha}}'\tilde{\boldsymbol{\alpha}})$ and the residual sum of squares is $\|\tilde{\mathbf{r}}\|^2 - (\tilde{\mathbf{r}}'\tilde{\boldsymbol{\alpha}})^2/\|\tilde{\boldsymbol{\alpha}}\|^2$, while the total sum of squares is $\|\tilde{\mathbf{r}}\|^2$. The Coefficient of Determination ("R squared") is, in expectation, equal to:

$$R^2 = \frac{(\tilde{\mathbf{r}}'\tilde{\boldsymbol{\alpha}})^2}{\|\tilde{\boldsymbol{\alpha}}\|^2 \|\tilde{\mathbf{r}}\|^2} = (\mathrm{IC})^2 \tag{4.57}$$

And we can link the coefficient of determination in predictive regressions to the Information Ratio:

$$\mathrm{IR} = \sqrt{R^2 n} \tag{4.58}$$

If there are T investment periods in a year, the annualized Information has a convenient form as a function of per-period cross-sectional R squared

$$\mathrm{IR} = \sqrt{R^2 n T} = \mathrm{IC}\sqrt{nT} \tag{4.59}$$

Otherwise stated, the annualized Information Ratio is equal to the Information Coefficient times the *independent number of return forecasts in a year.*

Modeling: Metrics

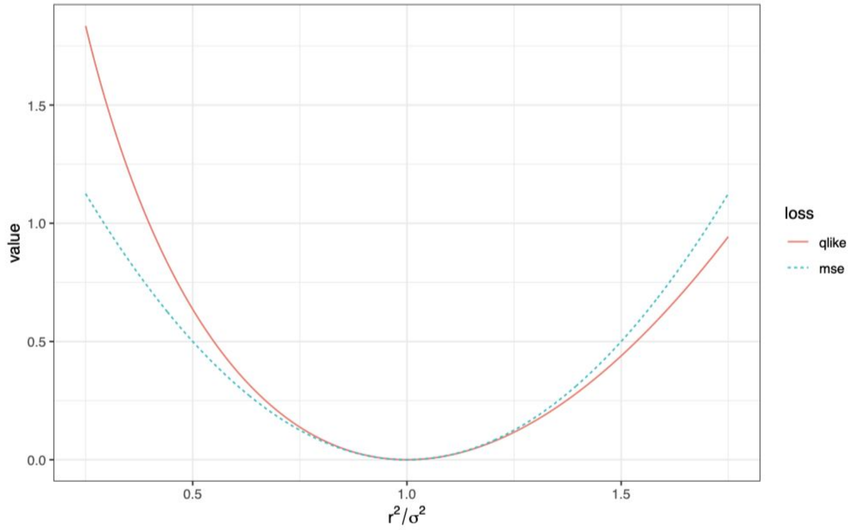
- A preliminary observation:
    a. **Any non-fiction book contains a hidden epistemology book in itself:** how do you form, extend, simplify and update your beliefs? **Every statement about actions is a statement about methods**
    b. And this applies to in-person interactions, from intellectual debates to first blind dates, to this course
    c. The problem is that placing that hidden book in plain sight is difficult, and who would read it anyway?
- In this context, there are two chapters in the book:
    a. What metrics do we use to rank and select models?
    b. How do we generalize?
- Today: first question. For another day: backtesting

Evaluating Factor Models

- R^2 on total returns.
  - Is it good? Discuss
- R^2 on idio returns.
  - You are assessing the quality of alpha. Is it good?
- Volatility estimates:
  - Robust when using volatility proxies

$$\mathrm{QLIKE}(\hat{\sigma}, r) := \frac{1}{n} \sum_{i=1}^{n} \left(\frac{r_i^2}{\hat{\sigma}_i^2} - \log\left(\frac{r_i^2}{\hat{\sigma}_i^2}\right) - 1 \right)$$

$$\mathrm{MSE}(\hat{\sigma}, r) := \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}_i^2 \left(\frac{r_i^2}{\hat{\sigma}_i^2} - 1 \right)^2$$



Fact: Not much different in practice

Additional Metrics

- Realized Variance of the minimum variance portfolio
  - Lower is better. Measures the performance of the precision metric
  - Check also the turnover
- Turnover of the covariance matrix.
  - Recommended: relative day-to-day change in Frobenius norm of the precision matrix
- Realized betas vs predicted on representative portfolios
  - Use existing strategy portfolios is possible
  - Alternatively, factor portfolios

Quiz (send to gp2642@nyu.edu in 10 mins **max**)

1. What ***can't*** you do to a factor model?
   a. Add a factor loading that is the first factor's terms, squared
   b. Square the factor returns of the first factor
   c. Remove the first factors
2. Can you write a min volatility constraint?
   a. Yes, it's doable and tractable
   b. No
   c. Yes, but it's not convex
3. Exponential utility is the same as
   a. Risk neutrality of the investor
   b. Risk neutrality as wealth increases
   c. Constant Absolute Risk Aversion
4. The Sharpe Ratio provides a bound on probability of max loss
   a. Only if variance is finite
   b. Only if returns are gaussian
   c. Only if returns are bounded
5. The annualied Information Ratio is proportional to
   a. Square root of R squared
   b. Number of positions per period
   c. Number of investment periods per year
6. Everything else being equal, positive autocorrelated returns
   a. Result in lower Sharpe
   b. Result in higher returns and higher vol
   c. Result in lower vol