Lesson ~ 27

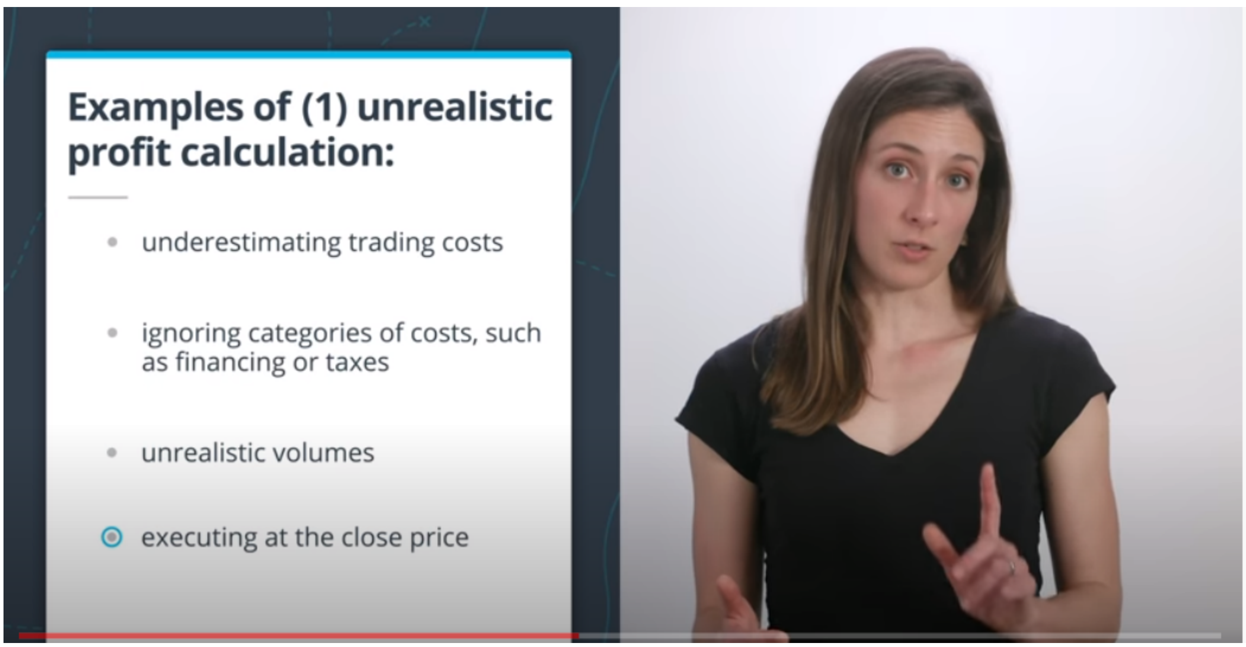
Files	
Tags	

25. Intro to backtesting

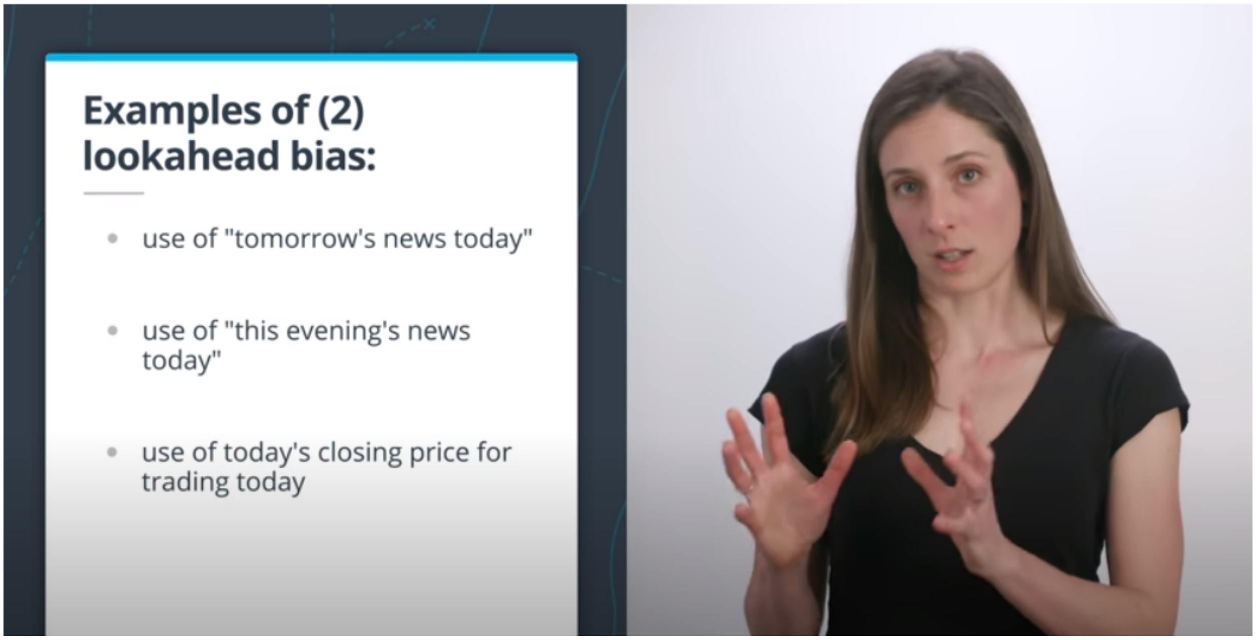
- Backtesting Validity

Valid backtest must satisfy 2 conditions

condition 1. Unrealistic profit calculation

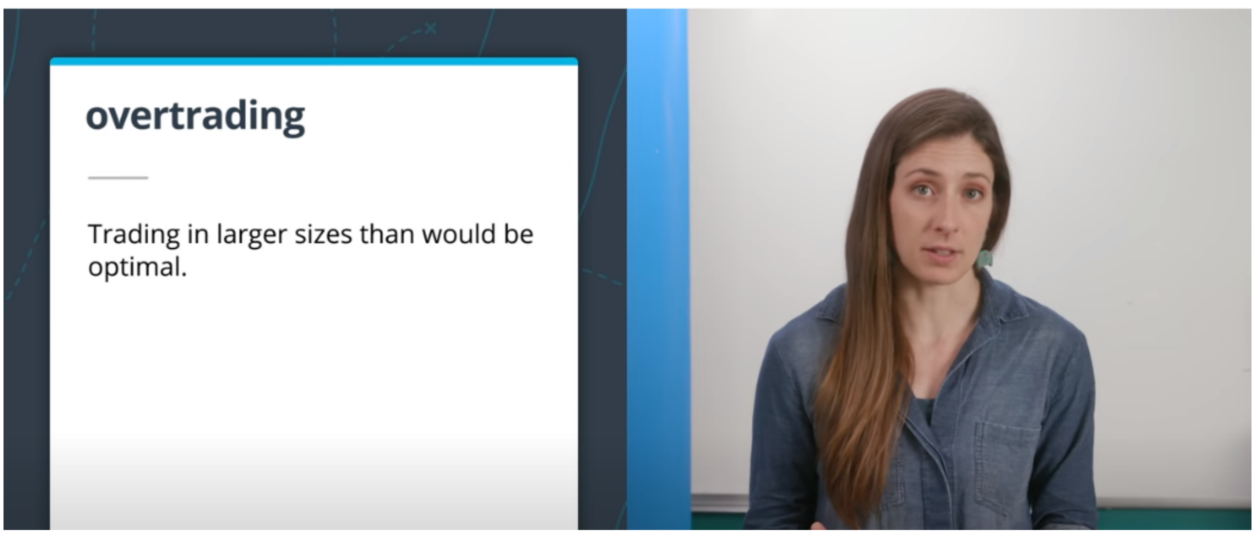


condition 2. lookahead bias



- Overtrading

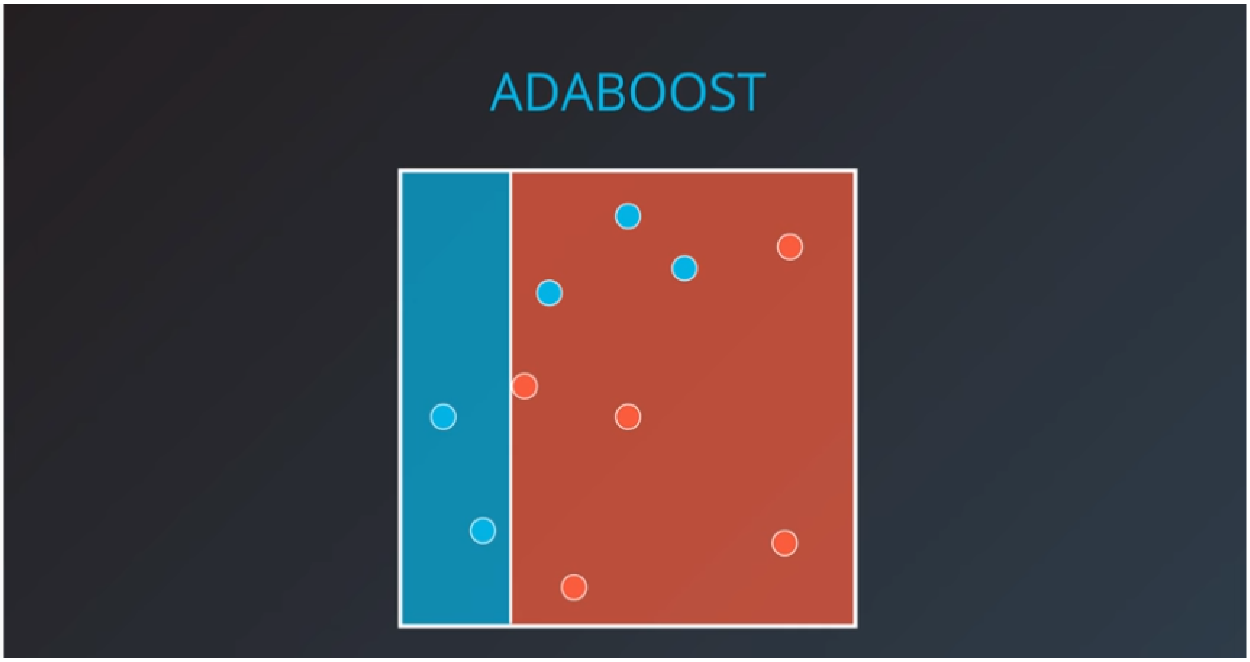
Overtrading can kill an entire strategy. Also overfitting can lead overtrading.

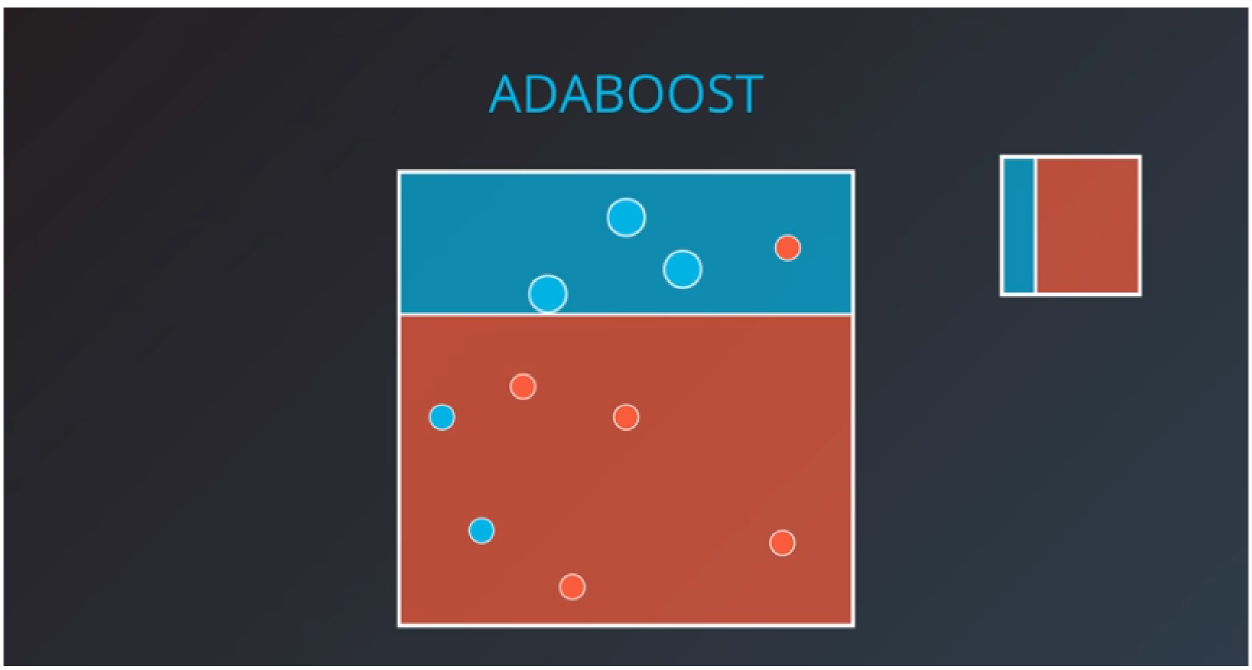


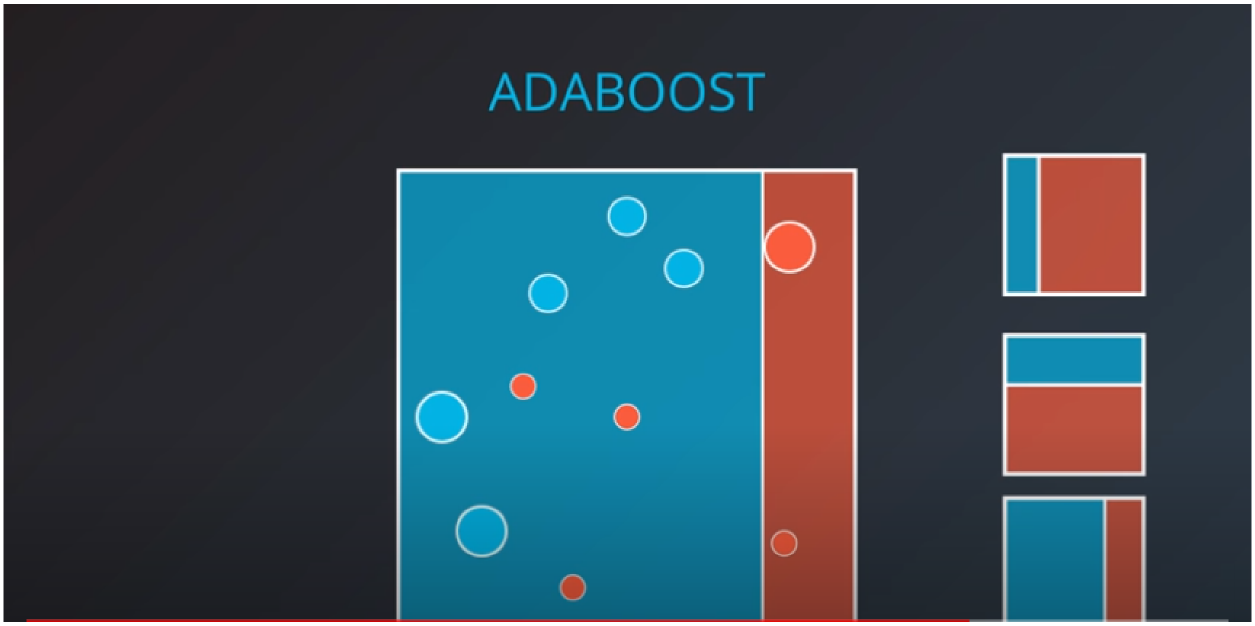
- Structure change

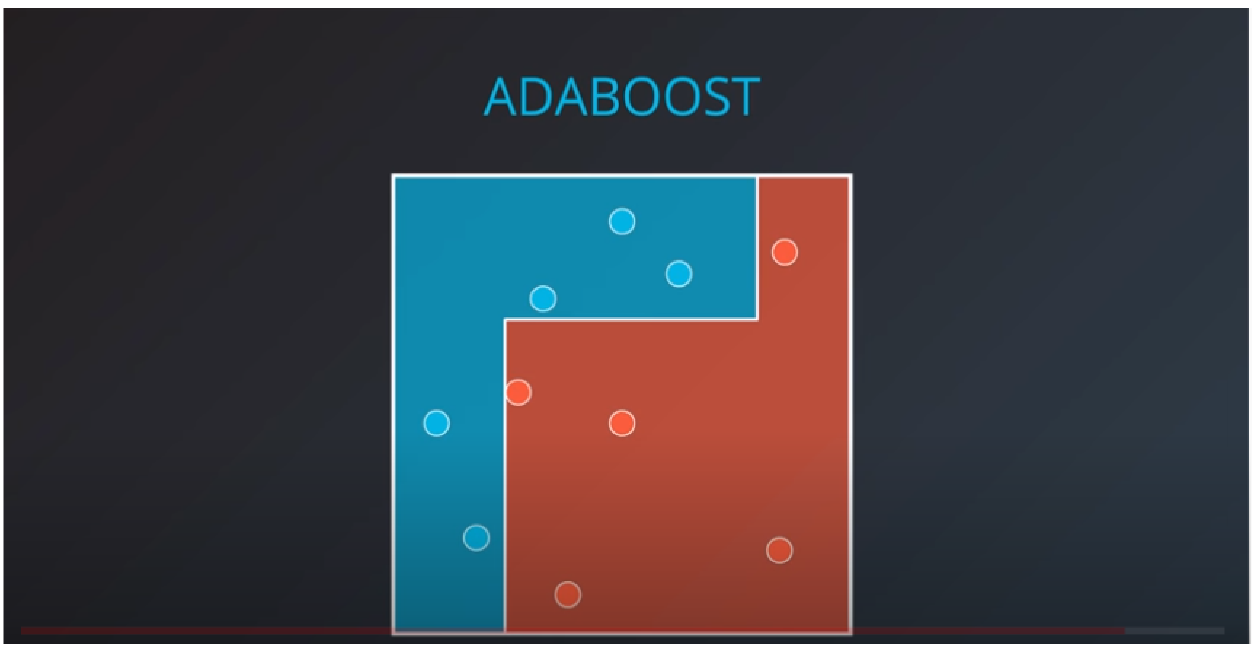
If the model's performance on the test set is low, there are 2 possible reason. One is overfitting. The other one is structural change which means market structure is different over the test set

- Gradient Boosting


ADABOOST

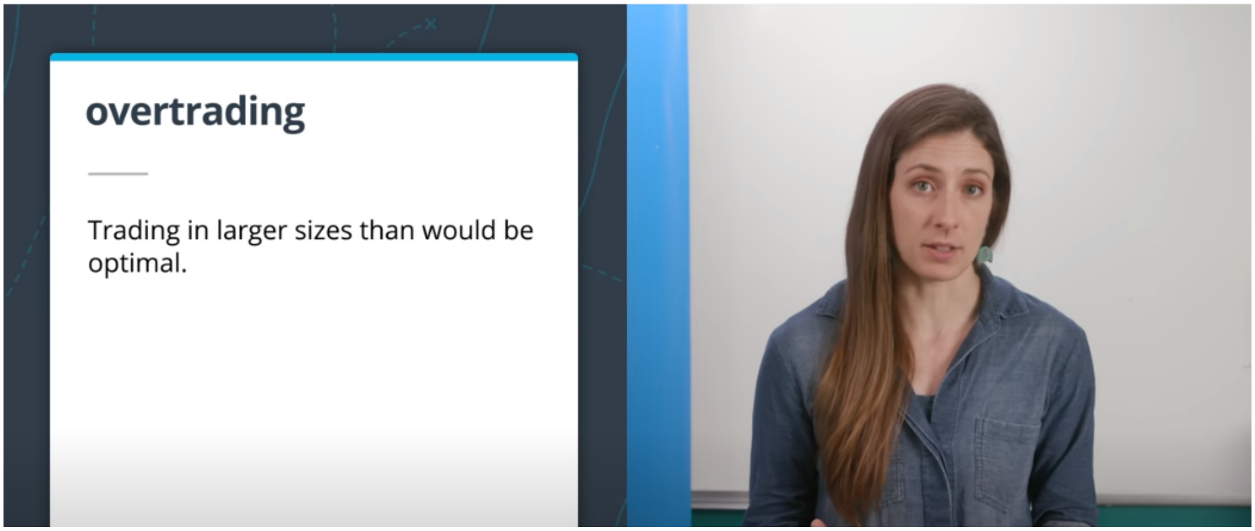

ADABOOST





details are later session

26. Optimization with Transaction Costs



시뮬레이션

- 데이터 가용성
- t기에 실재 매매를 한다 치면
- t-1기까지의 데이터를 사용해서 매매전략을 수립하고
- t+1기에 그 성과를 확인할 수 있음
- Barra에서는 이걸 시뮬레이션 해서 쓸 수 있도록 해주고 있음

포트폴리오 이론 vs 현실 백테스팅

몇가지 전제가 다름

- 이론에서는 W를 사용. W는 합이 1이 되는 숫자. 백테스팅에서는 실제 달러를 가지고 실험
- 백테스팅에서는 나의 거래가 시장에 영향을 줄 수 있다.
- 그러므로 그 영향의 크기도 거래에 반영해야 할 수 있음
- 거래비용 (슬리피지, 수수료)

최적화 문제

- 제약식 없는 최적화

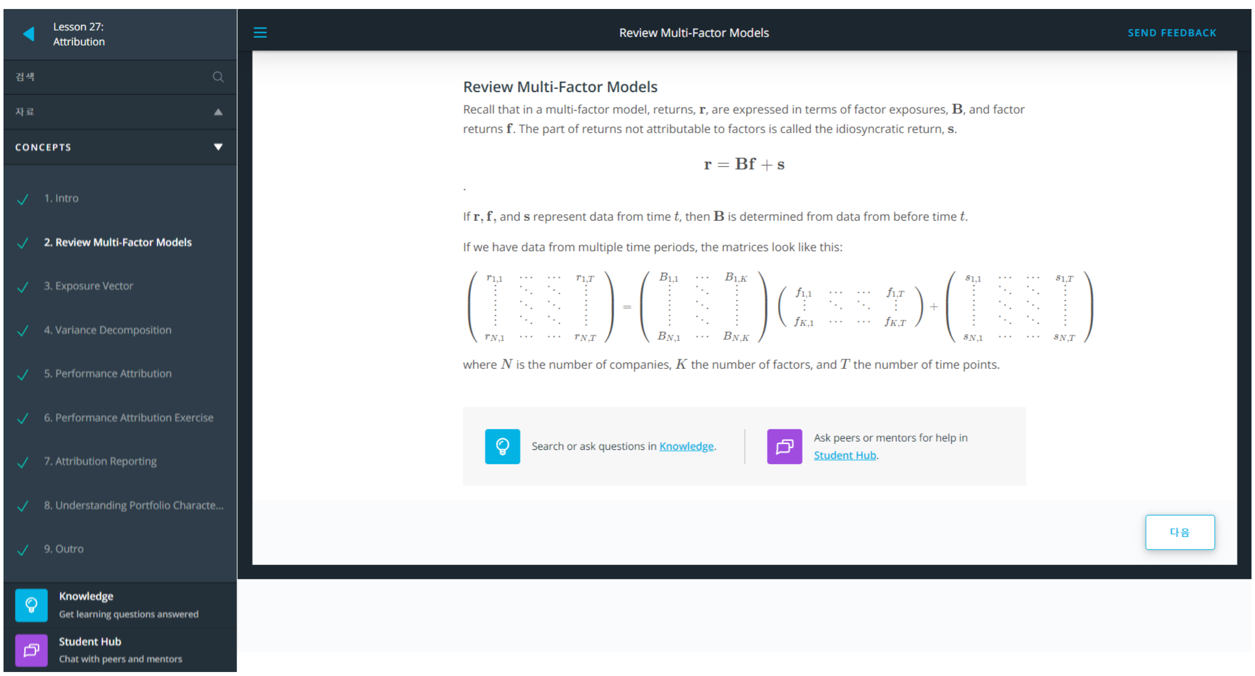
- 제약식 있는 최적화 (e.g.VaR, 최소분산, 거래비용)

그래서 실제로 쓰나요? → 쓴다고 하네요

27. Attribution

1. Intro

2. Review Multi-Factor Model



r : return, B : factor exposures, f : factor return, s : 특이성 return

r,f,s : 시간 t에서의 data

B : 시간 t이전에 정해진 data

3. Exposure Vector

$$\mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ \vdots \\ \vdots \\ h_N \end{pmatrix}$$

holding in company 1

$$\mathbf{B}_1 = \begin{array}{c} \begin{pmatrix} B_{1,1} \\ \vdots \\ \vdots \\ \vdots \\ B_{N,1} \end{pmatrix} \end{array} \begin{array}{c} *h_1 \\ *h_2 \\ \vdots \\ *h_N \end{array}$$

$$B_{1,1} * h_1 + \ldots + B_{N,1} * h_N = \mathbf{B}_1 \cdot \mathbf{h} = E_1$$

portfolio’s exposure to factor 1

holding vector : 각 자산에서 얼만큼의 돈을 각 회사에 투자 해야 하는가를 설명하는 요소

B : 각 회사에서 노출된 요소

E : the exposure vector of the portfolio

$$\left.\begin{matrix} E_1 \\ E_2 \\ E_3 \end{matrix}\right\} \quad \mathbf{E} = \begin{pmatrix} E_1 \\ \vdots \\ \vdots \\ \vdots \\ E_K \end{pmatrix}$$

$$\begin{pmatrix} B_{1,1} & \cdots & \cdots & \cdots & B_{1,N} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ B_{K,1} & \cdots & \cdots & \cdots & B_{K,N} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ \vdots \\ \vdots \\ h_N \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1 \cdot \mathbf{h} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{B}_K \cdot \mathbf{h} \end{pmatrix} = \begin{pmatrix} E_1 \\ \vdots \\ \vdots \\ \vdots \\ E_N \end{pmatrix} = \mathbf{E}$$

결과적으로 위와 같은 방식으로 E를 구할 수 있음

4. Variance Decomposition

1. 이전 단계에서 배운 covariance matrix를 분석해 봅시당

$$\mathrm{Var}[\mathbf{r}] = \boldsymbol{\Sigma} = \overbrace{\mathbf{B}^T\mathbf{F}\mathbf{B} + \mathbf{S}}^{\text{covariance matrix}}$$

2. 앞 뒤로 holding vector 곱함 : variance of our portfolio

$$\mathbf{h}^T\boldsymbol{\Sigma}\mathbf{h} = \mathbf{h}^T(\mathbf{B}^T\mathbf{F}\mathbf{B} + \mathbf{S})\mathbf{h}$$

$$\mathbf{h}^T\boldsymbol{\Sigma}\mathbf{h} = \underbrace{\mathbf{h}^T(\mathbf{B}^T}_{\text{exposure vector} \longrightarrow \mathbf{E}^T}\mathbf{F}\mathbf{B})\mathbf{h} + \mathbf{h}^T\mathbf{S}\mathbf{h}$$

$$\mathbf{h}^T\boldsymbol{\Sigma}\mathbf{h} = \underbrace{\mathbf{E}^T\mathbf{F}\mathbf{E}}_{\text{factor contributions}} + \underbrace{\mathbf{h}^T\mathbf{S}\mathbf{h}}_{\text{idiosyncratic contribution}}$$

3. 총 합 값으로 나눠버림 : variance decomposition

$$\frac{\mathbf{h}^T\mathbf{\Sigma}\mathbf{h}}{\mathbf{h}^T\mathbf{\Sigma}\mathbf{h}} = 1 = \frac{\mathbf{E}^T\mathbf{F}\mathbf{E}}{\mathbf{h}^T\mathbf{\Sigma}\mathbf{h}} + \frac{\mathbf{h}^T\mathbf{S}\mathbf{h}}{\mathbf{h}^T\mathbf{\Sigma}\mathbf{h}}$$

$$1 = \sum_{i=1}^{K} \underbrace{E_i \frac{(\mathbf{F}\mathbf{E})_i}{\mathbf{h}^T\mathbf{\Sigma}\mathbf{h}}}_{\text{i-th factor contribution}} + \underbrace{\frac{\mathbf{h}^T\mathbf{S}\mathbf{h}}{\mathbf{h}^T\mathbf{\Sigma}\mathbf{h}}}_{\text{idiosyncratic}}$$

i-th factor contribution : hedge 혹은 다른 요소들에 영향을 받은 경우 음수가 될 수 있음

idosyncratic : 음수가 될 수 없음

5. Performance Attribution

P&L(Profit and Loss) 분해

$$\begin{pmatrix} h_1 & \dots & \dots & \dots & h_N \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ \vdots \\ \vdots \\ r_N \end{pmatrix} = \mathbf{h}^T\mathbf{r}$$

$$\mathbf{r} = \mathbf{B}\mathbf{f} + \mathbf{s}$$

$$\mathbf{h}^T\mathbf{r} = \mathbf{h}^T\mathbf{B}\mathbf{f} + \mathbf{h}^T\mathbf{s}$$

$$\mathbf{h}^T\mathbf{r} = \underbrace{\mathbf{h}^T\mathbf{B}}_{\mathbf{E}^T}\mathbf{f} + \mathbf{h}^T\mathbf{s}$$

exposure vector $\longrightarrow \mathbf{E}^T$

$$\mathbf{h}^T\mathbf{r} = \overbrace{\sum_i E_i f_i}^{\text{includes both alpha and risk factors}} + \underbrace{\mathbf{h}^T\mathbf{s}}_{\text{P\&L unexplained by alpha or risk factors}}$$

$h^T s$: 특이성 요소, 각각 stock 선택에 의해 설명 되며 alpha 혹은 risk factor로 설명할 수 없음

$\sum_i E_i f_i$: alpha, risk factor를 포함한 요소, 예를들어 momentum factor가 큰 경우 P&L은 momentum에 의해 설명될 수 있음

6. Performance Attribution Exercise

7. Attribution Reporting

Attribution Reporting

Let's take this opportunity to look at an example attribution report in a fund's documentation.

In this first example, we can see the risk of a portfolio decomposed using a fundamental risk model. Fundamental factors are factors based on common sources of risk. Their meaning remains the same over time, even if the factor exposures are updated daily. The total predicted active risk of 3.63% annual volatility can be split into a specific/idiosyncratic component, which accounts for 39% of variance (as calculated using a variance decomposition), and a factor component, which can be further split into the contributions of 3 fundamental factors.

	Pred Risk (% Ann)	% of Variance
Total Active Risk	3.63%	100%
Specific Active Risk	2.25%	39%
Factor Active Risk (Fund)	2.84%	61%
Style (Fund)	1.95%	34%
Industry (Fund)	1.77%	29%
Market (Fund)	0.25%	-1%

You can do the same sort of attribution with statistical risk factors, but the individual risk factors are hard to interpret. In the example below, most of the risk is attributed to Statistical Factors 2, 1 and 6, however this does not immediately provide a lot of insight. Additional analysis would seek to understand whether other, interpretable factors are similar to Statistical Factor 6.

	Factor Exposure	Pred Risk (% Ann)	% of Variance
Total Active Risk		4.42%	100%
Specific Active Risk		2.48%	31%
Factor Active Risk		3.66%	69%
Statistical Factor 2	0.0191%	2.30%	27.0%
Statistical Factor 1	-0.0132%	1.55%	12.3%
Statistical Factor 6	-0.0128%	1.42%	10.3%
Statistical Factor 10	0.0096%	1.08%	5.91%
Statistical Factor 8	-0.0091%	1.07%	5.91%
Statistical Factor 13	0.0070%	0.80%	3.30%

8. Understanding Portfolio Character

Understanding Portfolio Characteristics

There are a few other things we can calculate that help us understand our portfolio's performance. For each time period:

- GMV (gross market value) is the sum of the absolute value of your holdings, long and short. This is a good gauge of the overall size of your portfolio. GMV = $\sum_i |h_i|$.
- Net holdings tell you the relative balance of long to short positions. Net holdings = $\sum_i h_i$.
- You can also calculate the total long and short holdings. Total long = $\sum_{h_i>0} h_i$, total short = $\sum_{h_i<0} h_i$.
- Total dollars traded tells you the approximate value of the trades you made. It can help you get a sense of how much trading you're doing, and the value of your trades relative to your holdings. Dollars traded = $\sum_i |h_{i,t} - h_{i,t-1}|$.

- GMV : holding들의 합산
- long : holding 들 중 양수(가격 상승 요소)의 합산
- short : holding 들 중 음수(가격 하락 요소)의 합산
- 변화량의 합산은 얼마나 거래를 했는지를 확인할 수 있음 유동성..?

9. Outro