

Risk modelling in equities: theory vs practice

Evgeny Rafikov
Evgeny Vasin

January 2014, Perm, Russia

О чем этот доклад

Движение акций \iff риск

Виды движений:

- Систематические, действующие на группу акций
- Специфические для данной акции

Систематические движения — способ отличать факты от слов

О чем ЭТОТ доклад

Движение акций \iff риск

Виды движений:

- Систематические, действующие на группу акций
- Специфические для данной акции

Систематические движения — способ отличать факты от слов

Setup

Вселенная: N акций, где N достаточно велико

Примеры: STOXX (600 европейских акций), TOPX (1000 японских акций), Russell 3000 (3000 американских акций), ...

Дневной горизонт

Одна география

Setup

Вселенная: N акций, где N достаточно велико

Примеры: STOXX (600 европейских акций), TOPX (1000 японских акций), Russell 3000 (3000 американских акций), ...

Дневной горизонт

Одна география

Setup

Вселенная: N акций, где N достаточно велико

Примеры: STOXX (600 европейских акций), TOPX (1000 японских акций), Russell 3000 (3000 американских акций), ...

Дневной горизонт

Одна география

Риск: systematic and stock-specific

Приближение 1: весь риск — специфический, все движения акций независимы в линейном смысле:

$$R_i(t) = \varepsilon_i(t)$$

Обозначения

$R_i(t)$: return on stock i over period t

$\varepsilon_i(t)$: случайная величина

"Российский фондовый рынок пошел вверх на 1.5%"

Возможно, Приближение 1 не совсем верно.

Риск: systematic and stock-specific

Приближение 1: весь риск — специфический, все движения акций независимы в линейном смысле:

$$R_i(t) = \varepsilon_i(t)$$

Обозначения

$R_i(t)$: return on stock i over period t

$\varepsilon_i(t)$: случайная величина

"Российский фондовый рынок пошел вверх на 1.5%"

Возможно, Приближение 1 не совсем верно.

Риск: systematic and stock-specific

Приближение 1: весь риск — специфический, все движения акций независимы в линейном смысле:

$$R_i(t) = \varepsilon_i(t)$$

Обозначения

$R_i(t)$: return on stock i over period t

$\varepsilon_i(t)$: случайная величина

"Российский фондовый рынок пошел вверх на 1.5%"

Возможно, Приближение 1 не совсем верно.

Риск: systematic and stock-specific

Приближение 2: движение акции = движение рынка +
движение акции по сравнению с рынком

$$R_i(t) = F_m(t) + \varepsilon_i(t)$$

Обозначения

$F_m(t)$: return of the market factor over period t

Если прогнать регрессию, окажется, что этот фактор
существенный.

Риск: systematic and stock-specific

Приближение 2: движение акции = движение рынка +
движение акции по сравнению с рынком

$$R_i(t) = F_m(t) + \varepsilon_i(t)$$

Обозначения

$F_m(t)$: return of the market factor over period t

Если прогнать регрессию, окажется, что этот фактор
существенный.

Риск: systematic and stock-specific

Приближение 3: а что если добавить сектор?

$$R_i(t) = F_m(t) + F_s(t) + \varepsilon_i(t)$$

Приближение 4: а что если добавить страну?

$$R_i(t) = F_m(t) + F_s(t) + F_c(t) + \varepsilon_i(t)$$

Приближение 5, Приближение 6, ...

Риск: systematic and stock-specific

Приближение 3: а что если добавить сектор?

$$R_i(t) = F_m(t) + F_s(t) + \varepsilon_i(t)$$

Приближение 4: а что если добавить страну?

$$R_i(t) = F_m(t) + F_s(t) + F_c(t) + \varepsilon_i(t)$$

Приближение 5, Приближение 6, ...

Риск: systematic and stock-specific

Приближение 3: а что если добавить сектор?

$$R_i(t) = F_m(t) + F_s(t) + \varepsilon_i(t)$$

Приближение 4: а что если добавить страну?

$$R_i(t) = F_m(t) + F_s(t) + F_c(t) + \varepsilon_i(t)$$

Приближение 5, Приближение 6, ...

Риск: systematic and stock-specific

Движение акций \iff риск

Результат: движения рынка разделяются на

- систематические, действующие на группу акций
- специфические для данной акции

$$R_i(t) = F_m(t) + F_s(t) + F_c(t) + \varepsilon_i(t)$$

Планы на ближайший час

- Научимся строить Fama-French style риск-модели
- Узнаем об основных факторах
- Посмотрим, что можно делать с помощью моделей
- Посмотрим, какие могут быть проблемы при построении
- Посмотрим на чисто статистические модели

- Научимся строить Fama-French style риск-модели
- Узнаем об основных факторах
- Научимся оценивать модели
- Посмотрим, что можно делать с помощью моделей
- Посмотрим, какие могут быть проблемы при построении
- Посмотрим на чисто статистические модели

Общий вид рассматриваемых моделей

$$R_i(t) = \sum L_{i,k}(t-1)F_k(t) + \varepsilon_i(t)$$

Известно:

- $L_{i,k}(t-1)$: exposure of stock i to factor k at the end of $(t-1)$
- $R_i(t)$: return of stock i over period t

Оценивается:

- $F_k(t)$: return on unit exposure to factor k over period t
- $\varepsilon_i(t)$: residual return on stock i over period t

Физический смысл factor loadings

Значения $L_{i,k}(t)$ определяют, насколько сильно акция i подвержена фактору k

Сделаем нормировку (для всех кроме dummies):

- Выбросим outliers

- $$\overline{L_i(t)} = \frac{(L_{i,k}(t) - E(L_{i,k}(t)))}{\sigma(L_{i,k}(t))}$$

Физический смысл $\overline{L_{i,k}(t)} = 1$: акция с таким показателем в момент времени t отличается от средней акции в нашей вселенной на одно стандартное отклонение

Физический смысл factor loadings

Значения $L_{i,k}(t)$ определяют, насколько сильно акция i подвержена фактору k

Сделаем нормировку (для всех кроме dummies):

- Выбросим outliers

- $$\overline{L_i(t)} = \frac{(L_{i,k}(t) - E(L_{i,k}(t)))}{\sigma(L_{i,k}(t))}$$

Физический смысл $\overline{L_{i,k}(t)} = 1$: акция с таким показателем в момент времени t отличается от средней акции в нашей вселенной на одно стандартное отклонение

Физический смысл factor returns

Физический смысл $F_k(t)$: насколько сильно тот или иной фактор в исследуемом периоде повлиял на returns.

Физический смысл $F_k(t) = 0.01$: акция с единичным exposure к этому фактору, в среднем, имела return на 1% больший, чем средняя акция на рынке.

Регрессионная процедура

Проблема: если в модели есть Market и либо Sector dummies, либо Country dummies, у нас есть мультиколлинеарность:

Market	Country1	Country2	Sector1	Sector2
1	1	0	1	0
1	1	0	0	1
1	0	1	0	1
1	1	0	0	1
1	0	1	1	0

Регрессионная процедура

Решение — ограничения на dummies.

Например, если есть страны и сектора, можно ввести:

$$\sum F_s^k = 0$$

$$\sum F_c^k = 0$$

Тогда return сектора/страны — это excess return по сравнению с рынком

Проблема: small caps будут давать слишком большой шум

Регрессионная процедура

Решение — ограничения на dummies.

Например, если есть страны и сектора, можно ввести:

$$\sum F_s^k = 0$$

$$\sum F_c^k = 0$$

Тогда return сектора/страны — это excess return по сравнению с рынком

Проблема: small caps будут давать слишком большой шум

Регрессионная процедура

Как бороться с шумом?

Ввести веса!

Две идеи:

- Рынок - это совокупность всех акций. Значит, веса стоит давать в зависимости от рыночной капитализации (обычно какой-то функции от нее).
- Доходности - это зашумленные наблюдения "истинных параметров". Значит, взвешивать нужно с использованием какой-то функции от волатильности.

Регрессионная процедура

Как бороться с шумом?

Ввести веса!

Две идеи:

- Рынок - это совокупность всех акций. Значит, веса стоит давать в зависимости от рыночной капитализации (обычно какой-то функции от нее).
- Доходности - это зашумленные наблюдения "истинных параметров". Значит, взвешивать нужно с использованием какой-то функции от волатильности.

Итоговый вариант: Generic Risk Model

$$R_{t,i} = F_t \cdot L_{t,i} + \varepsilon_{i,t}$$

$(1 \times K) \quad (K \times 1)$

Матричная форма факторной модели:

$$R = F \cdot L + E$$

$(T \times N) \quad (T \times K) \quad (K \times N) \quad (T \times N)$

N – количество акций.

T – количество временных промежутков.

K – количество факторов.

Примеры значений.

$$N = 2000, \quad T = 500, \quad K = 100$$

Итоговый вариант: Generic Risk Model

$$R_{t,i} = F_t \cdot L_{t,i} + \varepsilon_{i,t}$$

$(1 \times K) \quad (K \times 1)$

Матричная форма факторной модели:

$$R = F \cdot L + E$$

$(T \times N) \quad (T \times K) \quad (K \times N) \quad (T \times N)$

N – количество акций.

T – количество временных промежутков.

K – количество факторов.

Примеры значений.

$$N = 2000, \quad T = 500, \quad K = 100$$

Итоговый вариант: Generic Risk Model

$$R_{t,i} = F_t \cdot L_{t,i} + \varepsilon_{i,t}$$

$(1 \times K) \quad (K \times 1)$

Матричная форма факторной модели:

$$R = F \cdot L + E$$

$(T \times N) \quad (T \times K) \quad (K \times N) \quad (T \times N)$

N – количество акций.

T – количество временных промежутков.

K – количество факторов.

Примеры значений.

$$N = 2000, \quad T = 500, \quad K = 100$$

- Научимся строить Fama-French style риск-модели
- Узнаем об основных факторах
- Научимся оценивать модели
- Посмотрим, что можно делать с помощью моделей
- Посмотрим, какие могут быть проблемы при построении
- Посмотрим на чисто статистические модели

Факторы на примере Barra

- Market
- Sector
- Country
- Currency
- Price momentum
- Volatility
- Turnover
- Size
- Value
- Growth
- Dividend Yield
- Earnings Yield
- Leverage

Factor: Sector

Нужен, чтобы захватывать движение сектора

Простой вариант: sector dummy в какой-нибудь классификации, например GICS L1 (10) /L2 (24)

Альтернатива: собирать сектора из более детальной классификации (GICS L3: 68, L4: 154)

Factor: Country

Нужен, чтобы захватывать движение страны

Простой вариант, country dummy.

1 если компания зарегистрирована в стране, 0 если не зарегистрирована.

Альтернатива: фундаментальный фактор, основанный на деятельности компании в разных странах

Style factor MOMENTUM

$$w_1 \cdot \text{HWALF} + w_2 \cdot \text{MOM}_{12,1} + w_3 \cdot \text{MOM}_{6,1}$$

HWALF Historical weekly alpha. Intercept, α_i of a regression of weekly asset returns $R_{i,t}$, against weekly returns of the cap-weighted estimation universe R_{Mt}

$$R_{i,t} = \alpha_i + \beta_i R_{Mt} + \varepsilon_{i,t}$$

E.g. use 2 years of weekly data (104 weeks) and exponential weighting with a half-life of 52 weeks.

Style factor MOMENTUM

$$w_1 \cdot \text{HWALF} + w_2 \cdot \text{MOM}_{12,1} + w_3 \cdot \text{MOM}_{6,1}$$

$\text{MOM}_{12,1}$ 12-month relative strength, lagged by 1 month.

$$\text{MOM}_{12,1} = \sum_{t=T-\text{NP}}^{T-1} \ln(1 + R_{i,t}) - \sum_{t=T-\text{NP}}^{T-1} \ln(1 + R_{0,t})$$

$R_{i,t}$: monthly asset return at time t

$R_{0,t}$: monthly risk-free return at time t

NP : number of periods in sum, here NP = 12

$\text{MOM}_{6,1}$ 6-month relative strength, lagged by 1 month. Defined as $\text{MOM}_{12,1}$ but uses NP = 6

Style factor VOLATILITY

$$w_1 \cdot \text{HWBT} + w_2 \cdot \text{CMR}_{12,0} + w_3 \cdot \text{DSTD}_{65,23}$$

HWBT Historical weekly beta. Slope β_i of the regression for **HWALF**

CMR_{12,0} Cumulative range, defined as

$$\text{CMRA} = \ln(1 + \max(Z_t)) - \ln(1 + \min(Z_t))$$

$$Z_t = \sum_{s=1}^t \ln(1 + R_{i,s}) - \sum_{s=1}^t \ln(1 + R_{0,s})$$

$R_{i,s}$: monthly asset return at time s

$R_{0,s}$: monthly risk-free return at time s

Style factor VOLATILITY

$$w_1 \cdot \text{HWBT} + w_2 \cdot \text{CMR}_{12,0} + w_3 \cdot \text{DSTD}$$

DSTD Daily asset volatility. Exponentially-weighted standard deviation of local daily asset returns. E.g.

DSTD = **DSTD**_{65,23} Uses 65 daily returns and a half-life of 23 days.

$$\text{DSTD}_{65,23} = \left(\sum_{s=t-64}^t \lambda^{t-s} \cdot (R_{i,s} - \bar{R}_i)^2 / \sum_{s=t-64}^t \lambda^{t-s} \right)^{1/2}$$

$$\lambda = 0.5^{1/\tau}, \text{ and } \tau = 23 \text{ days}$$

Style factor LIQUIDITY

$$w_1 \cdot \text{LSTOA} + w_2 \cdot \text{LSTOQ} + w_3 \cdot \text{LSTOM}$$

LSTOA Log of annual share turnover; an indicator of the average liquidity of an asset over 1 year

$$\text{LSTOA} = \ln\left(\frac{1}{\text{NP}} \cdot \sum_{t=T-\text{NP}+1}^T \frac{\text{VOL}_t}{\text{NOS}_t}\right)$$

VOL_t : monthly volume of shares traded at time t

NOS_t : number of shares outstanding at time t

NP : number of periods in sum, e.g. $\text{NP} = 12$

LSTOQ Log of quarterly share turnover; i.e. **LSTOA** with $\text{NP} = 3$

LSTOM Log of monthly share turnover; i.e. **LSTOA** with $\text{NP} = 1$

Style factor SIZE

$$w_1 \cdot \text{LNCAP} + w_2 \cdot \text{ASSI}$$

LNCAP Log of the month-end issuer capitalization

ASSI Log of total assets; an indicator of fundamental firm size

Style factor VALUE

$$w_1 \cdot \text{BTP} + w_2 \cdot \text{STP}$$

- BTP** Book-to-price ratio. The last published book value of common equity divided by the current issuer capitalization.
- STP** Sales-to-price ratio. Sales over the last 12 months divided by the current issuer capitalization

Style factor GROWTH

$$w_1 \cdot \text{AGRO} + w_2 \cdot \text{SAGRO} + w_3 \cdot \text{EGRO} + w_4 \cdot \text{EGRSF} + w_5 \cdot \text{EGRLF}$$

AGRO Trailing growth of total assets. Slope coefficient of the reported total assets regressed against time over the last 5 years.

SAGRO Trailing growth of annual sales. Slope coefficient of the reported annual sales regressed against time over the last 5 years.

EGRO Trailing growth of annual net earnings. Slope coefficient of the reported annual net earnings regressed against time over the last 5 years

Style factor GROWTH

$$w_1 \cdot \text{AGRO} + w_2 \cdot \text{SAGRO} + w_3 \cdot \text{EGRO} + w_4 \cdot \text{EGRSF} + w_5 \cdot \text{EGRLF}$$

EGRSF Short-term predicted earnings growth, defined as:

$$\text{EGRSF} = \frac{\text{EPS}_{12\text{F}} - \text{EPS}_{12\text{B}}}{\text{abs}(\text{EPS}_{12\text{B}})}$$

$\text{EPS}_{12\text{F}}$: forward-looking earnings per share

$\text{EPS}_{12\text{B}}$: backward-looking earnings per share, defined as

$$\text{EPS}_{12\text{B}} = (\text{M} \cdot \text{EPS}_0 + (12 - \text{M}) \cdot \text{EPS}_1) / 12$$

EPS_0 : last reported earnings per share.

EPS_1 : consensus of EPS forecasts for current fiscal year.

M : number of months since the end of last fiscal year

EGRLF Long-term (3-5 years) predicted earnings growth, consensus of analyst estimates

Style factor DIVIDEND YIELD

$$1.0 \cdot \text{YLD}$$

YLD Annualized dividend per share divided by the current price

Style factor EARNINGS YIELD

$$w_1 \cdot \text{ETP} + w_2 \cdot \text{CETP} + w_3 \cdot \text{ROE} + w_4 \cdot \text{PETP}$$

ETP Trailing earnings-to-price ratio. Net earnings over the last 12 months divided by the current issuer capitalization.

CETP Cash earnings-to-price ratio. Cash earnings over the last 12 months divided by the current issuer capitalization.

ROE Return on equity. Net earnings over the last 12 months divided by the current issuer capitalization.

PETP Predicted earnings-to-price ratio. 12-month forward-looking earnings per share divided by current price. Forward-looking earnings per share are defined as

$$\text{EPS}_{12F} = (M \cdot \text{EPS}_1 + (12 - M) \cdot \text{EPS}_2) / 12$$

EPS_1 : consensus of EPS forecasts for current fiscal year.

EPS_2 : consensus of EPS forecasts for next fiscal year.

M : number of months remaining to the end of current fiscal year

Style factor LEVERAGE

$$w_1 \cdot \text{BLEV} + w_2 \cdot \text{MLEV}$$

BLEV Book leverage. Computed as

$$\text{BLEV} = \frac{\text{BV} + \text{PF} + \text{LD}}{\text{BV}}$$

BV : most recent book value of common equity

PF : most recent book value of preferred equity

LD : most recent book value of long-term debt

MLEV Market leverage. Defined in the same way but BV is replaced by the most recent issuer capitalization.

Что еще можно придумать?

- Price reversion
- Analyst sentiment
- Analyst revisions
- ...

Factor: Price reversion

Компании, которые дорожали на протяжении последних нескольких дней, подешевеют. Компании, которые дешевели, будут подорожают.

Пример: return over 3 days

Улучшение: считать residual returns reversion, а не просто returns reversion

Улучшение: смотреть еще и на trade imbalance

...

Factor: Price reversion

Компании, которые дорожали на протяжении последних нескольких дней, подешевеют. Компании, которые дешевели, будут подорожают.

Пример: return over 3 days

Улучшение: считать residual returns reversion, а не просто returns reversion

Улучшение: смотреть еще и на trade imbalance

...

Factor: Price reversion

Компании, которые дорожали на протяжении последних нескольких дней, подешевеют. Компании, которые дешевели, будут подорожают.

Пример: return over 3 days

Улучшение: считать residual returns reversion, а не просто returns reversion

Улучшение: смотреть еще и на trade imbalance

...

Factor: Price reversion

Компании, которые дорожали на протяжении последних нескольких дней, подешевеют. Компании, которые дешевели, будут подорожают.

Пример: return over 3 days

Улучшение: считать residual returns reversion, а не просто returns reversion

Улучшение: смотреть еще и на trade imbalance

...

Analyst sentiment

В поведении акций, которые нравятся аналитикам, есть что-то общее по отношению к акциям, которые аналитикам не нравятся

Пример: посчитать средний рейтинг (buy/sell/neutral) по всем аналитикам на рынке

Улучшение: взвесить хороших и плохих аналитиков с разными весами

Улучшение: учитывать разное время обновления прогнозов

...

Analyst sentiment

В поведении акций, которые нравятся аналитикам, есть что-то общее по отношению к акциям, которые аналитикам не нравятся

Пример: посчитать средний рейтинг (buy/sell/neutral) по всем аналитикам на рынке

Улучшение: взвесить хороших и плохих аналитиков с разными весами

Улучшение: учитывать разное время обновления прогнозов

...

Analyst sentiment

В поведении акций, которые нравятся аналитикам, есть что-то общее по отношению к акциям, которые аналитикам не нравятся

Пример: посчитать средний рейтинг (buy/sell/neutral) по всем аналитикам на рынке

Улучшение: взвесить хороших и плохих аналитиков с разными весами

Улучшение: учитывать разное время обновления прогнозов

...

Analyst sentiment

В поведении акций, которые нравятся аналитикам, есть что-то общее по отношению к акциям, которые аналитикам не нравятся

Пример: посчитать средний рейтинг (buy/sell/neutral) по всем аналитикам на рынке

Улучшение: взвесить хороших и плохих аналитиков с разными весами

Улучшение: учитывать разное время обновления прогнозов

...

Analyst revisions

Рынок реагирует на факт изменения предсказаний аналитиков

Пример: сумма изменений предсказаний EPS FY1 за какой-то период

Улучшение: использовать 12MF

Улучшение: кластеры предсказаний (по времени)

...

Analyst revisions

Рынок реагирует на факт изменения предсказаний аналитиков

Пример: сумма изменений предсказаний EPS FY1 за какой-то период

Улучшение: использовать 12MF

Улучшение: кластеры предсказаний (по времени)

...

Analyst revisions

Рынок реагирует на факт изменения предсказаний аналитиков

Пример: сумма изменений предсказаний EPS FY1 за какой-то период

Улучшение: использовать 12MF

Улучшение: кластеры предсказаний (по времени)

...

Analyst revisions

Рынок реагирует на факт изменения предсказаний аналитиков

Пример: сумма изменений предсказаний EPS FY1 за какой-то период

Улучшение: использовать 12MF

Улучшение: кластеры предсказаний (по времени)

...

Analyst revisions

Рынок реагирует на факт изменения предсказаний аналитиков

Пример: сумма изменений предсказаний EPS FY1 за какой-то период

Улучшение: использовать 12MF

Улучшение: кластеры предсказаний (по времени)

...

- Научимся строить Fama-French style риск-модели
- Узнаем об основных факторах
- Научимся оценивать модели
- Посмотрим, что можно делать с помощью моделей
- Посмотрим, какие могут быть проблемы при построении
- Посмотрим на чисто статистические модели

Схема оценки параметров модели на примере Barra

Cross-sectional regression with weights proportional to $\sqrt{\text{mcap}}$

Outlier stock returns detected and filtered with a **two-step** regression scheme:

- (1)
$$R_i = \beta_{iM} \cdot F_M + \sum_C \beta_{iC} \cdot F_C + \sum_J \beta_{iJ} \cdot F_J + \sum_S \beta_{iS} \cdot F_S + \varepsilon_i$$
- (2) identify outlier component e_i : set $\varepsilon_{\max} = 3 \cdot \text{std}_i(\varepsilon_i)$
if $\text{abs}(\varepsilon_i) > \varepsilon_{\max}$ then set $e_i = \varepsilon_i - \text{sgn}(\varepsilon_i) \cdot \varepsilon_{\max}$, otherwise
 $e_i = 0$
- (3)
$$R_i - e_i = \beta_{iM} \cdot \tilde{F}_M + \sum_C \beta_{iC} \cdot \tilde{F}_C + \sum_J \beta_{iJ} \cdot F_J + \sum_S \beta_{iS} \cdot \tilde{F}_S + \varepsilon_i$$

- Научимся строить Fama-French style риск-модели
- Узнаем об основных факторах
- Научимся оценивать модели
- Посмотрим, что можно делать с помощью моделей
- Посмотрим, какие могут быть проблемы при построении
- Посмотрим на чисто статистические модели

Что можно делать с моделью

- Проанализировать портфель (статика)
- Захеджировать портфель (статика)
- Проанализировать стратегию (динамика)
- Проанализировать и улучшить discretionary стратегию (динамика)
- Построить факторную стратегию (динамика)

Анализ портфеля: factor exposures

Обозначения

$L_{P,k}(t)$: portfolio exposure to factor k at the end of period t

В относительных значениях:

$$L_{P,k}(t) = \sum L_{i,k}(t)w_i(t)$$

Физический смысл: на сколько стандартных отклонений факторный профиль портфеля отличается от рыночного портфеля

В абсолютных значениях:

$$L_{P,k}(t) = \sum L_{i,k}(t)N_i(t)P_i(t)$$

Анализ портфеля: factor exposures

Обозначения

$L_{P,k}(t)$: portfolio exposure to factor k at the end of period t

В относительных значениях:

$$L_{P,k}(t) = \sum L_{i,k}(t)w_i(t)$$

Физический смысл: на сколько стандартных отклонений факторный профиль портфеля отличается от рыночного портфеля

В абсолютных значениях:

$$L_{P,k}(t) = \sum L_{i,k}(t)N_i(t)P_i(t)$$

Анализ портфеля: factor risks

Обозначения

$Q_{P,k}(t)$: portfolio risk по отношению к фактору k в конце периода t

$$Q_{P,k}(t) = \sum L_{i,k} w_i \sigma(F_k)$$

В абсолютных значениях:

$$Q_{P,k}(t) = \sum L_{i,k}(t) N_i(t) P_i(t) \sigma(F_k)$$

Физический смысл: на сколько подешевеет/подорожает портфель, если фактор j пойдет вверх на одно стандартное отклонение

Анализ портфеля: factor risks

Обозначения

$Q_{P,k}(t)$: portfolio risk по отношению к фактору k в конце периода t

$$Q_{P,k}(t) = \sum L_{i,k} w_i \sigma(F_k)$$

В абсолютных значениях:

$$Q_{P,k}(t) = \sum L_{i,k}(t) N_i(t) P_i(t) \sigma(F_k)$$

Физический смысл: на сколько подешевеет/подорожает портфель, если фактор j пойдет вверх на одно стандартное отклонение

Анализ портфеля: portfolio risk

Сделаем два предположения:

- $\text{corr}(R_k, \varepsilon_i) = 0$
- $\text{corr}(\varepsilon_i, \varepsilon_j) = 0$

Тогда:

$$\begin{aligned}\text{var}(R_P) &= \text{var}_{\text{systematic}} + \text{var}_{\text{specific}} \\ \text{var}_{\text{systematic}} &= L_P \Sigma L_P' \\ \text{var}_{\text{specific}} &= W_P \text{var}(\varepsilon) W_P'\end{aligned}$$

Обозначения

Σ : factor variance-covariance matrix

Формирование factor tracking portfolio

Задача: построить портфель с $L^k = 1$ и $L^{j \neq k} = 0$

Решение

1. Вспомним, что

$$L_{P,k} = \sum L_{i,k} w_i$$

2. Добавим:

$$\sum w_i = 0 \quad \sum |w_i| = 1$$

Можно получить много таких портфелей

Не все из них будут оптимальными с точки зрения residual risk и не все будут торгуемыми

Формирование factor tracking portfolio

Задача: построить портфель с $L^k = 1$ и $L^{j \neq k} = 0$

Решение

1. Вспомним, что

$$L_{P,k} = \sum L_{i,k} w_i$$

2. Добавим:

$$\sum w_i = 0 \quad \sum |w_i| = 1$$

Можно получить много таких портфелей

Не все из них будут оптимальными с точки зрения residual risk и не все будут торгуемыми

Формирование factor tracking portfolio

Задача: построить портфель с $L^k = 1$ и $L^{j \neq k} = 0$

Решение

1. Вспомним, что

$$L_{P,k} = \sum L_{i,k} w_i$$

2. Добавим:

$$\sum w_i = 0 \quad \sum |w_i| = 1$$

Можно получить много таких портфелей

Не все из них будут оптимальными с точки зрения residual risk и не все будут торгуемыми

Формирование factor tracking portfolio

Альтернатива: оптимизационная задача

$$\begin{aligned} Q'_P &\rightarrow \min \\ \text{s.t. } L_{P,k} &= 1 \end{aligned}$$

Смысл: минимизируем риск, сохраняя константный exposure к желаемому фактору.

Задача аналогична задаче максимизации exposure к одному фактору при сохранении константного риска.

Формирование factor tracking portfolio

Альтернатива: оптимизационная задача

$$\begin{aligned} Q'_P &\rightarrow \min \\ \text{s.t. } L_{P,k} &= 1 \end{aligned}$$

Смысл: минимизируем риск, сохраняя константный exposure к желаемому фактору.

Задача аналогична задаче максимизации exposure к одному фактору при сохранении константного риска.

Анализ стратегии: PnL attribution

Дано: есть некоторая стратегия (т.е. последовательность портфелей по всех точках некоторого периода времени).

Также есть последовательность, доходность стратегии во всех точках этого периода.

Задача: понять, за счет чего портфель демонстрировал ту или иную доходность.

Анализ стратегии (PnL attribution)

Вспомним для single stock:

$$R_i(t) = R_{\text{systematic}}(t) + R_{\text{specific}}(t)$$

Для портфеля аналогично:

$$R_P(t) = R_{\text{systematic}}(t) + R_{\text{specific}}(t)$$

Анализ стратегии (PnL attribution)

Вспомним для single stock:

$$R_i(t) = R_{\text{systematic}}(t) + R_{\text{specific}}(t)$$

Для портфеля аналогично:

$$R_P(t) = R_{\text{systematic}}(t) + R_{\text{specific}}(t)$$

Анализ стратегии (PnL attribution)

$$R_P(t) = R_{\text{systematic}}(t) + R_{\text{specific}}(t)$$

$$R_{\text{systematic}}(t) = \sum L_{i,k}(t-1)F_k(t)$$

$$R_{\text{specific}}(t) = \sum w_i F_i(t)$$

Результат: мы можем понять систематическую компоненту PnL стратегии и stock picking компоненту

Анализ стратегии (PnL attribution)

$$R_P(t) = R_{\text{systematic}}(t) + R_{\text{specific}}(t)$$

$$R_{\text{systematic}}(t) = \sum L_{i,k}(t-1)F_k(t)$$

$$R_{\text{specific}}(t) = \sum w_i F_i(t)$$

Результат: мы можем понять систематическую компоненту PnL стратегии и stock picking компоненту

Создание факторной стратегии

Задача: сделать такую стратегию, которая бы максимально точно (по крайней мере априори) отслеживала доходность фактора или факторов.

Основное отличие от factor tracking portfolio: ненулевые издержки на изменение портфеля.

Создание факторной стратегии

Варианты, как можно формулировать задачу:

- Лимитировать, какая часть портфеля может меняться при каждой ребалансировке
Проблема: transaction cost для разных акций будет очень разным
- Лимитировать максимальную ожидаемую стоимость ребалансировки
Проблема: формула transaction costs может быть нетривиальной
- Включать стоимость ребалансировки в оптимизируемую формулу

- Научимся строить Fama-French style риск-модели
- Узнаем об основных факторах
- Научимся оценивать модели
- Посмотрим, что можно делать с помощью моделей
- Посмотрим на чисто статистические модели
- Посмотрим, какие могут быть проблемы при построении

Stepwise optimization

Sample covariance matrix

$$Q = (R - ER)^t \times (R - ER)/T, \quad Q = (q_{i,j})_{(N \times N)}$$

First iteration step: find 1st factor loadings.

$$\begin{matrix} L_1^t & \times & Q & \times & L_1 & \longrightarrow & \max \\ (1 \times N) & & (N \times N) & & (N \times 1) & & (L_1) \end{matrix}$$

$$\text{subj to: } \begin{cases} L_1^t \times L_1 = 1 \\ L_{1,1} > 0 \end{cases}$$

As a result get first factor returns (scores):

$$\begin{matrix} F_1 & = & R & \times & L_1 & / & N \\ (T \times 1) & & (T \times N) & & (N \times 1) & & \end{matrix}$$

Stepwise optimization, k-th step

k-th optimization step, $k = 1, \dots, K$, where K is the number of factors

$$\underset{(1 \times N)}{\mathbf{L}_k^t} \times \underset{(N \times N)}{\mathbf{Q}} \times \underset{(N \times 1)}{\mathbf{L}_k} \longrightarrow \max_{(\mathbf{L}_k)}$$

$$\text{subj to: } \begin{cases} \mathbf{L}_k^t \times \mathbf{L}_k = 1 \\ \mathbf{L}_{k,1} > 0 \\ \mathbf{F}_1^t \times \mathbf{F}_k = 0 \\ \dots \\ \mathbf{F}_{k-1}^t \times \mathbf{F}_k = 0 \end{cases}$$

As a result get first factor returns (scores):

$$\underset{(T \times 1)}{\mathbf{F}_k} = \underset{(T \times N)}{\mathbf{R}} \times \underset{(N \times 1)}{\mathbf{L}_k} / N$$

Maximum Likelihood Estimation

Assumptions: $\{\varepsilon_{i,t}\}$ are i.i.d. normal; $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$.

\mathbf{R} , matrix of returns is de-meaned. In the case where the factor returns \mathbf{F} are observed, the MLE estimate of the exposures \mathbf{B} , and specific risks $\{\sigma_i^2\}$ are calculated as

$$\mathbf{L} = (\mathbf{F}^t \cdot \mathbf{F})^{-1} \cdot \mathbf{Z}^t \cdot \mathbf{R}$$

$$\sigma^2 = \text{diag}(\mathbf{R}^t \times \mathbf{R}) / \mathbf{N} = \text{diag}(\mathbf{R}^t \times \mathbf{E})$$

No assumptions on distributional form of the observed factor returns \mathbf{F} are made.

Expectation-Maximization Estimation

Both factor returns \mathbf{F} and factor exposures \mathbf{L} are estimated through some iterative procedure. The reason is to choose a set of orthogonal factors that are optimal in some sense, given some prior info about distributional form of factor returns \mathbf{F} . As above we also assume that $\varepsilon_{t,i} \sim N(0, \sigma_i^2)$. We then have

$$(R_i | \mathbf{F}, L_i, \sigma_i^2) \sim N(\mathbf{F} \cdot L_i, \sigma_i^2),$$

Sufficient statistics for the above complete-data MLE estimators:

$$C_{\mathbf{F}\mathbf{F}} = \mathbf{F}^t \cdot \mathbf{F} / N, \quad \text{quad}C_{\mathbf{F}\mathbf{R}} = \mathbf{F}^t \cdot \mathbf{R} / N, \quad C_{\mathbf{R}\mathbf{R}} = \mathbf{R}^t \cdot \mathbf{R} / N.$$

We adopt an assumption for \mathbf{F} as a multivariate normal prior.

Expectation-Maximization Estimation

The conditional expectation of the sufficient statistics, given the observed data \mathbf{R} and current estimate of the parameters $(\mathbf{L}_i, \sigma_i^2)$ can be calculated:

$$E[C_{FF} | \mathbf{R}, \mathbf{L}, \sigma^2, \mathbf{Q}] = \mathbf{F}^t \cdot \mathbf{F} / N + \Delta$$

$$E[C_{FR} | \mathbf{R}, \mathbf{L}, \sigma^2, \mathbf{Q}] = \mathbf{F}^t \cdot \mathbf{R} / N$$

$$E[C_{RR} | \mathbf{R}, \mathbf{L}, \sigma^2, \mathbf{Q}] = \mathbf{R}^t \cdot \mathbf{R} / N$$

$$\text{where } \Delta = \mathbf{Q} - (\mathbf{Q} \cdot \mathbf{L})(\sigma^2 + \mathbf{L}^t \cdot \mathbf{Q} \cdot \mathbf{L})^{-1} \cdot (\mathbf{L}^t \cdot \mathbf{Q})$$

Expectation-Maximization Estimation

Expectation-step:

$$\begin{aligned}\Delta &= \Sigma - \Sigma \cdot L \cdot \delta \\ F &= R \cdot \delta\end{aligned}$$

where

$$\delta = \sigma^{-2} \cdot L^t \cdot \Sigma - \sigma^{-2} \cdot L^t \cdot (\Sigma^{-1} + L \cdot \sigma^{-2} \cdot L^t)^{-1} \cdot L \cdot \sigma^{-2} \cdot L^t \cdot \Sigma$$

Maximization-step:

$$\begin{aligned}L &= (F^t \cdot F + N \cdot \Delta)^{-1} \cdot F^t \cdot R \\ E &= R - F \cdot L \\ \sigma^2 &= \text{diag}(R^t \cdot E) / N = \text{diag}(E^t \cdot E + L^t \cdot \Delta \cdot L) \\ \Sigma &= F^t \cdot F / N + \Delta\end{aligned}$$

- Научимся строить Fama-French style риск-модели
- Узнаем об основных факторах
- Научимся оценивать модели
- Посмотрим, что можно делать с помощью моделей
- Посмотрим на чисто статистические модели
- Посмотрим, какие могут быть проблемы при построении

Основные проблемы и ошибки

Основная проблема: данные

Основная ошибка: forward looking bias

Данные: corporate actions

Corporate actions

- Stock split
- Дивиденды
- Rights
- ...и еще 80 разных специальных случаев

Влияют на данные для других факторов

Влияют на цены

Влияют на доходность стратегий (дивиденды!) и на бенчмарки

Данные: corporate actions

Corporate actions

- Stock split
- Дивиденды
- Rights
- ...и еще 80 разных специальных случаев

Влияют на данные для других факторов

Влияют на цены

Влияют на доходность стратегий (дивиденды!) и на бенчмарки

Данные: corporate actions

Corporate actions

- Stock split
- Дивиденды
- Rights
- ...и еще 80 разных специальных случаев

Влияют на данные для других факторов

Влияют на цены

Влияют на доходность стратегий (дивиденды!) и на бенчмарки

Данные: corporate actions

Corporate actions

- Stock split
- Дивиденды
- Rights
- ...и еще 80 разных специальных случаев

Влияют на данные для других факторов

Влияют на цены

Влияют на доходность стратегий (дивиденды!) и на бенчмарки

Данные: backfills

Backfills

- Особенно часто встречаются в данных по аналитикам
- Случаются в фундаментальных данных

Единственный правильный вариант: самим каждый день сохранять

Данные: backfills

Backfills

- Особенно часто встречаются в данных по аналитикам
- Случаются в фундаментальных данных

Единственный правильный вариант: самим каждый день сохранять

Ошибки: universe

Universe

- Неликвидные акции (в экстремальных случаях может мешать и оценке факторов)
- Short sell ограничения
- Forward looking bias: нельзя использовать будущие композиции вселенной

С чем надо сравнивать стратегии

Правильный ответ: со всем, с чем получится

- Индексы; для long only стратегии
- Universe: для long only стратегий
- Factor tracking portfolios для классических факторов

Горизонт backtest-a

Горизонт должен быть такой, чтобы:

- Было достаточно много перебалансировок
- Были разные периоды (кризис, рост, quant meltdown, Lehman crisis, QE, ...)
- Была видна динамика альфы

Risk Model Providers

MSCI Barra

Axioma

EM Applications

Most large investment banks