

Machine
Learning

Celebrity

FACTO
R

CTBC

投資名人堂



中信投資名人堂 組別 2-1

指導業師: 林裕訓

指導老師: 石百達、蔡芸琇

組員: 許惟傑(台大財金碩一)、林哲瑋(台大財金碩一)

徐慕華(台大國企碩二)、盧煒中(東吳巨資三年級)





01

投資名人介紹

02

研究方法

03

投資組合策略

04

結語

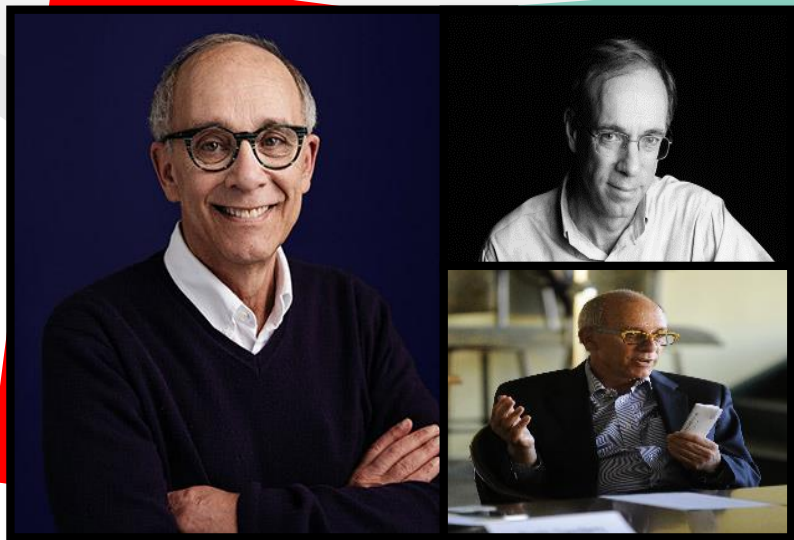
A light gray curved decorative line starts from the left edge of the slide, curves downwards and to the right, and then levels out towards the right edge.

First you imitate, then you innovate.

Miles Davis



投資名人介紹



由他領軍的孤松資本從最初募集800萬美元
目前已躍昇為管理230億美元的大型基金

史蒂芬·曼德爾 (Stephen Mandel)

Lone Pine Capital 避險基金經理人



投資哲學

曼德爾認為這樣的投資策略具長期有效性

01

非價值也非成長型投資人，尋找由
好人所經營的優質企業

02

方法:由下而上(bottom-up)
相信基於徹底了解企業之基本面能預判未來成長

03

偏愛企業

- 以自己的優勢保持營收成長創造價值的企業
- 管理優良的企業
- 有新的企業策略但尚未反映至價格上的企業
- 持有期間非長期，買賣憑自身有力且迅速的判斷

Why Him?



資料長度足夠(超過15年)

限制:由於其持有其他標的
(如call/put, 無資料來源)
使用與文獻相同方法, 只專注在股票的部分

股票報酬成長可觀
從2004~2019年成長了5倍

Rocket資料偶爾有誤
須排除outlier

持有股票幾乎為美國交易所上市



研究方法



樣本

孤松資本每月投資組合月報酬
(去除選擇權及國際股)

樣本期間

2005/1-2019/12

目標

機器學習方式
仿造其選股風格

樣本數量

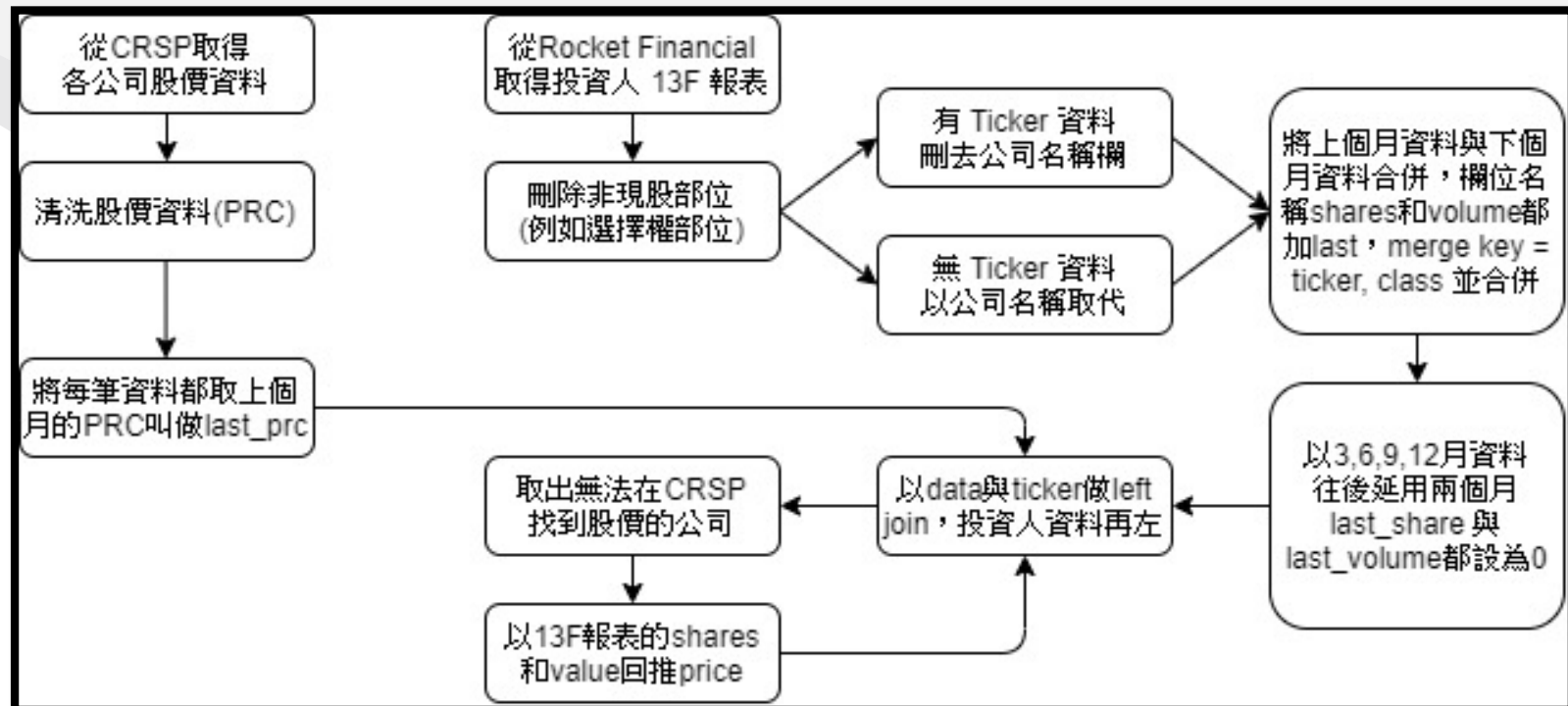
季與季之間持股不變

假設

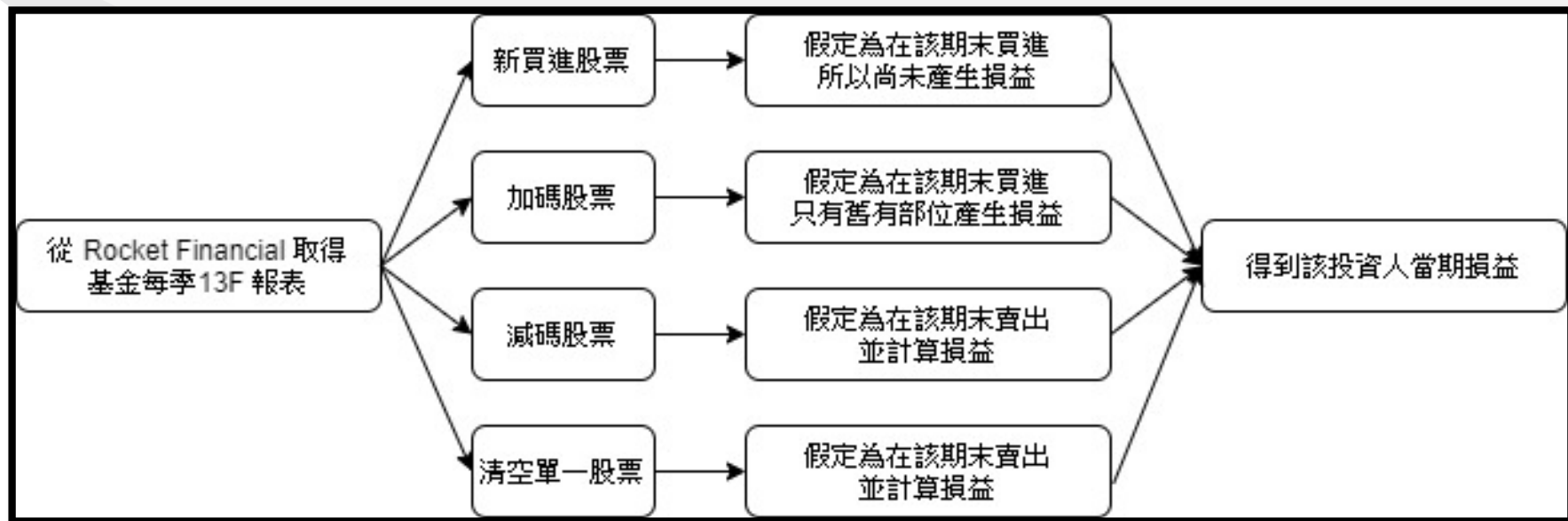
180筆

資料處理流程

利用爬蟲將Rocket投資人每月持股爬下來，進行資料處理



基金的資產會受到投資人贖回影響，所以我們進行以下處理得到投資人在當期的損益以計算報酬

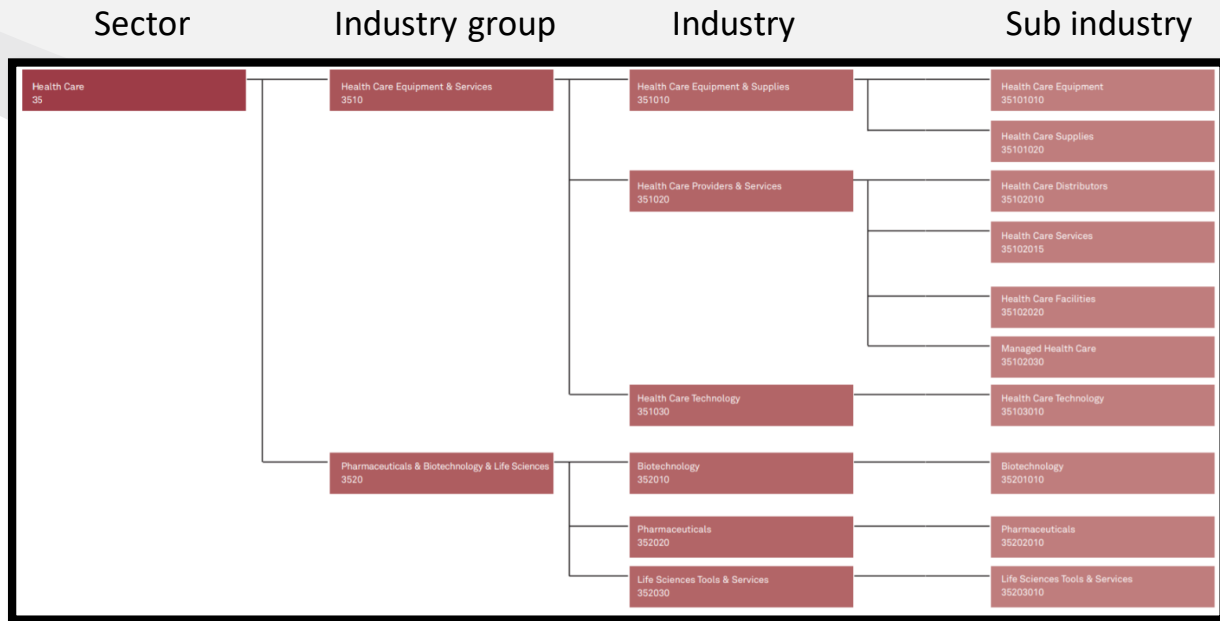


投資組合篩選概念



產業篩選

此處產業類別以GIC code 為標準。GIC code 是由MSCI 與 S&P 所編，目的為將各公司進行產業分類，其分類分為以下層級：11個 Sector，24 個 Industry group，69 個 Industry 和 158 個 Subindustry。我們採用第一個分級(Sector)為各個公司進行分類以避免分類過細



產業篩選

統計投資人每季股票
所屬產業

計算每個產業佔投組的比例
(以次數計算比例)



計算每個產業佔投組的比例
(以次數計算比例)

挑選產業累積占比大於等於80%
(若不滿五個產業，則補至五個)

因子篩選

以下為市場上比較常見的因子介紹:

- Fama-French提出之因子:

MKT: 市場風險溢酬

SMB3: 小型股投資組合報酬 - 大型股投資組合報酬 (以市值排序)

SMB5: SMB3的變形

HML: 高B/M股投資組合報酬 - 低B/M股投資組合報酬

UMD: 以前2-12個月報酬為基準，選出贏家與輸家組合贏家投資組合報酬 - 輸家投資組合報酬

RMW: 高營利投資組合報酬-低營利投資組合報酬

CMA: 高再投資投資組合報酬-低再投資組合報酬

- Andrea Frazzini提出之因子:

BAB: 低Beta投資組合報酬-高Beta投資組合報酬

QMJ: 高品質投資報酬報酬-低品質投資組合報酬

- 其他因子 (eg.景氣相關虛擬因子，股利相關因子)

因子模型選擇

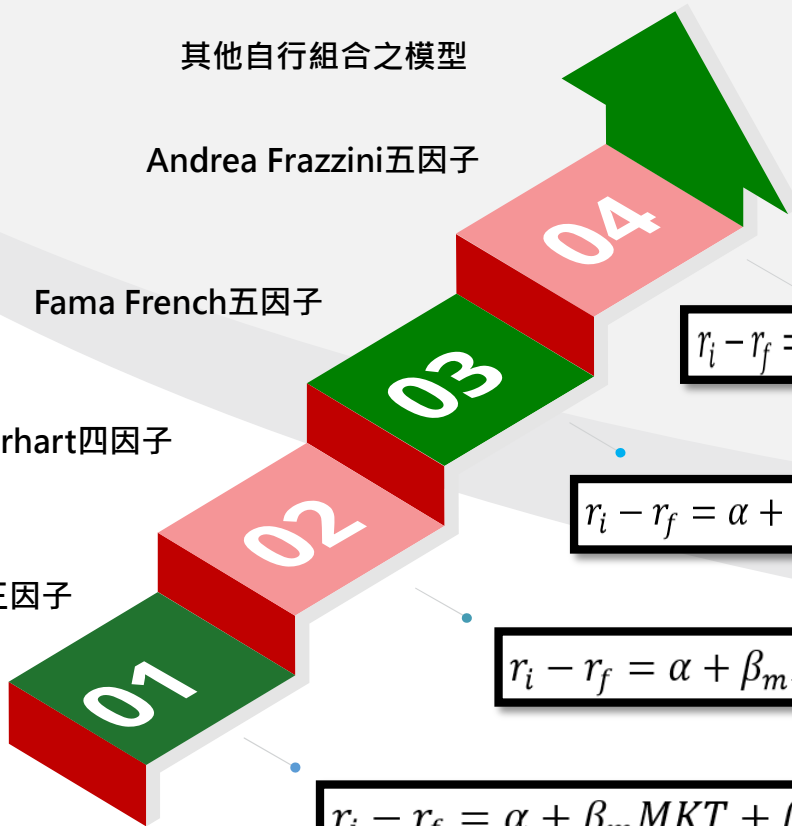
其他自行組合之模型

Andrea Frazzini五因子

Fama French五因子

Carhart四因子

Fama French三因子



$$r_i - r_f = \alpha + \beta_m MKT + \beta_s SMB + \beta_h HML + \beta_u UMD + \beta_q QMJ + \beta_b BAB + \varepsilon$$

$$r_i - r_f = \alpha + \beta_m MKT + \beta_s SMB + \beta_h HML + \beta_r RMW + \beta_c CMA + \varepsilon$$

$$r_i - r_f = \alpha + \beta_m MKT + \beta_s SMB + \beta_h HML + \beta_u UMD + \varepsilon$$

$$r_i - r_f = \alpha + \beta_m MKT + \beta_s SMB + \beta_h HML + \varepsilon$$

因子顯著性測試

Fama-French三因子模型

| | |
|---------------------|--------|
| F(3,176) | 116.73 |
| R ² | 0.67 |
| Adj. R ² | 0.66 |

| | Est. | S.E. | t val. | p |
|-------------|-------|------|--------|------|
| (Intercept) | 0.25 | 0.23 | 1.09 | 0.28 |
| MKT | 1.03 | 0.06 | 17.39 | 0.00 |
| SMB3 | 0.03 | 0.10 | 0.27 | 0.79 |
| HML | -0.50 | 0.09 | -5.62 | 0.00 |

Standard errors: OLS

Fama-French 三因子模型

三因子模型顯示 **MKT 與 HML 非常顯著**
顯示 Stephen Mandel 的表現受到市場關連強
並傾向買入低 B/M 的股票(HML係數為負)

Carhart四因子模型

| | | | | |
|----------------------|-------|------|--------|------|
| F(4,175) | 88.76 | | | |
| R ² | 0.67 | | | |
| Adj. R ² | 0.66 | | | |
| | Est. | S.E. | t val. | p |
| (Intercept) | 0.27 | 0.23 | 1.22 | 0.23 |
| MKT | 1.01 | 0.06 | 16.48 | 0.00 |
| SMB3 | 0.03 | 0.10 | 0.25 | 0.81 |
| HML | -0.55 | 0.10 | -5.80 | 0.00 |
| UMD | -0.09 | 0.06 | -1.51 | 0.13 |
| Standard errors: OLS | | | | |

Carhart 四因子模型

四因子模型顯示 **MKT 與 HML 依然非常顯著**
市場動能因子 UMD 不顯著
顯示出 **Stephen Mandel 的報酬並非由追逐上漲股票而來**

Fama-French五因子模型

| | | | | |
|----------------------|-------|------|--------|------|
| F(5,174) | 74.70 | | | |
| R ² | 0.68 | | | |
| Adj. R ² | 0.67 | | | |
| | Est. | S.E. | t val. | p |
| (Intercept) | 0.26 | 0.23 | 1.13 | 0.26 |
| MKT | 1.00 | 0.06 | 15.74 | 0.00 |
| SMB5 | 0.08 | 0.11 | 0.74 | 0.46 |
| HML | -0.33 | 0.11 | -3.13 | 0.00 |
| RMW | 0.13 | 0.17 | 0.77 | 0.45 |
| CMA | -0.55 | 0.19 | -2.93 | 0.00 |
| Standard errors: OLS | | | | |

Fama-French 五因子模型

五因子模型顯示 **MKT、HML 與 CMA 非常顯著**。**CMA 因子顯示出 Stephen Mandel 投資方式較為積極**

因子顯著性測試

Andrea Frazzini模型

四因子+ BAB+ QMJ

F(6,173) 59.47

R² 0.67

Adj. R² 0.66

| | Est. | S.E. | t val. | p |
|-------------|-------|------|--------|------|
| (Intercept) | 0.39 | 0.24 | 1.62 | 0.11 |
| MKT | 0.96 | 0.07 | 13.54 | 0.00 |
| SMB3 | -0.01 | 0.11 | -0.10 | 0.92 |
| HML | -0.58 | 0.10 | -5.92 | 0.00 |
| UMD | -0.06 | 0.06 | -0.91 | 0.36 |
| BAB | -0.06 | 0.10 | -0.66 | 0.51 |
| QMJ | -0.17 | 0.13 | -1.28 | 0.20 |

Standard errors: OLS

變形

五因子+ BAB+ QMJ

F(7,172) 55.11

R² 0.69

Adj. R² 0.68

| | Est. | S.E. | t val. | p |
|-------------|-------|------|--------|------|
| (Intercept) | 0.39 | 0.24 | 1.66 | 0.10 |
| MKT | 0.92 | 0.07 | 13.04 | 0.00 |
| SMB5 | 0.05 | 0.11 | 0.47 | 0.64 |
| HML | -0.40 | 0.11 | -3.61 | 0.00 |
| RMW | 0.34 | 0.20 | 1.76 | 0.08 |
| CMA | -0.54 | 0.19 | -2.88 | 0.00 |
| BAB | -0.05 | 0.09 | -0.60 | 0.55 |
| QMJ | -0.32 | 0.15 | -2.19 | 0.03 |

Standard errors: OLS

加入新因子 BAB 與 QMJ
結果顯示 MKT 與 HML 非常顯著，
另外 QMJ 為些微顯著且係數為負

我們認為 QMJ 的係數為負的原因為
Andrea Frazzini 較常投資科技股股票。
因為科技股通常投資較積極，且
科技股波動率較大，會使科技股被分
類為較不具價值的股票，所以在最後
導致導致 QMJ 的係數為負

加入虛擬變數因子

以虛擬因子方式，新增景氣相關因子
(1為相對景氣好;0為相對景氣壞)

X5Yaverag 虛擬變數

Lag一年之公債 2y10y 利差，>五年平均為0，
反之為1

DOWN:

標準普爾指數，>五年平均為1，反之為0

VIX:

恐慌指數，<30為1，反之為0

模型(第六行)

前面所出現過因子(SMB選擇SMB3)+市場時期
虛擬變數

從本表觀察可發現，不論使用任何模型，
MKT,HML,CMA都保持顯著。三者用於衡量市
場時期的虛擬變數皆不顯著，表示好與壞時期在
長期影響不顯著

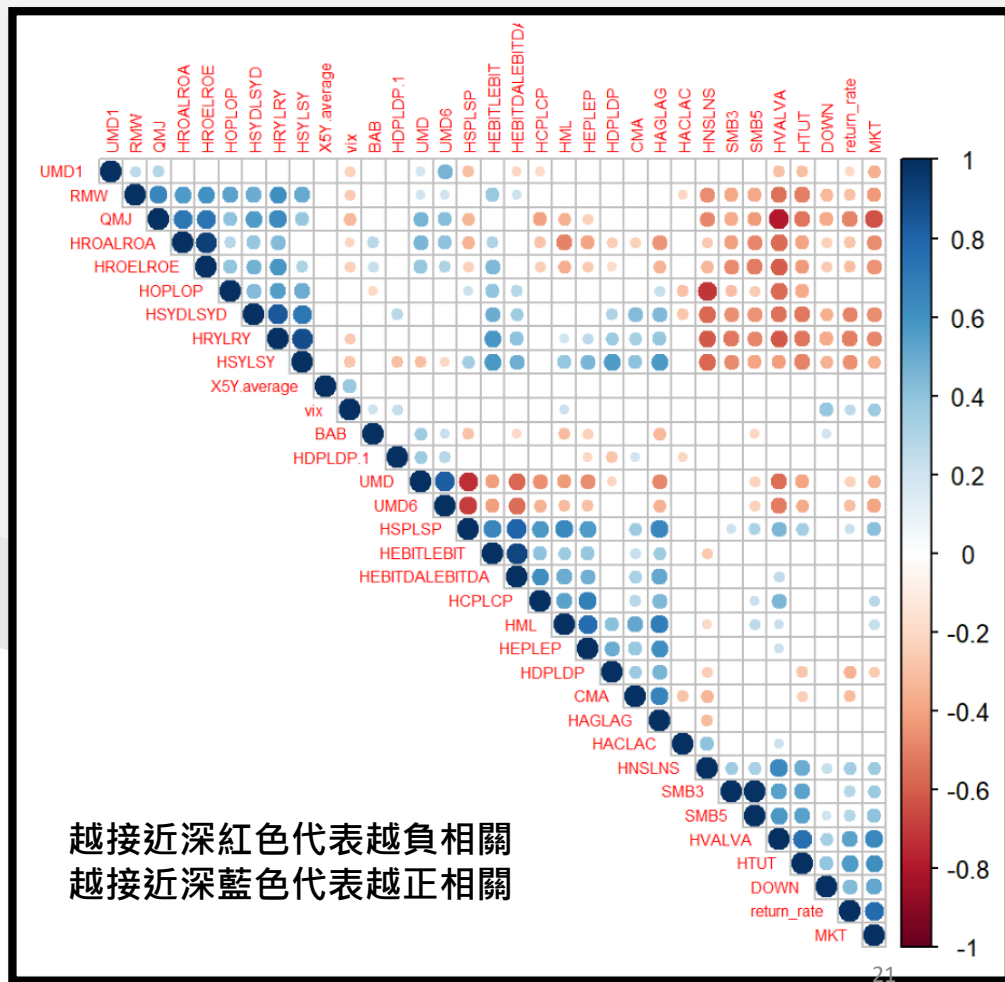
| | ThreeFactor | FourFactor | FiveFactor | BABQMJ1 | BABQMJ2 | WithTime |
|-------------|-----------------------|-----------------------|----------------------|-----------------------|-----------------------|-----------------------|
| (Intercept) | 0.245 (0.226) | 0.275 (0.226) | 0.258 (0.229) | 0.389 (0.240) | 0.392 (0.236) | -0.990 (1.064) |
| MKT | 1.031 *** (0.059) | 1.008 *** (0.061) | 0.995 *** (0.063) | 0.965 *** (0.071) | 0.923 *** (0.071) | 0.866 *** (0.080) |
| SMB3 | 0.028 (0.105) | 0.026 (0.104) | | -0.011 (0.108) | | 0.037 (0.109) |
| HML | -0.498 *** (0.088) | -0.553 *** (0.095) | -0.332 ** (0.106) | -0.585 *** (0.099) | -0.402 *** (0.111) | -0.445 *** (0.120) |
| UMD | | -0.086 (0.057) | | -0.056 (0.061) | | -0.077 (0.068) |
| SMB5 | | | 0.078 (0.106) | | 0.051 (0.107) | |
| RMW | | | 0.127 (0.166) | | 0.344 (0.196) | 0.304 (0.204) |
| CMA | | | -0.547 ** (0.187) | | -0.535 ** (0.186) | -0.548 ** (0.187) |
| BAB | | | | -0.064 (0.096) | -0.054 (0.091) | -0.085 (0.099) |
| QMJ | | | | -0.169 (0.132) | -0.322 * (0.147) | -0.228 (0.160) |
| X5Y.average | | | | | | -0.509 (0.512) |
| DOWN | | | | | | 0.674 (0.622) |
| vix | | | | | | 1.345 (1.167) |
| N | 180 | 180 | 180 | 180 | 180 | 180 |
| R2 | 0.666 | 0.670 | 0.682 | 0.673 | 0.692 | 0.698 |

*** p < 0.001; ** p < 0.01; * p < 0.05.

相關性分析

除了以上因子以外，我們也使用了額外的23個因子

在選擇前，我們先剔除與基本因子相關性過高的因子



最後留下相關性選擇之因子模型

HDPLDP (高股利殖利率)

高股利殖利率(最高10%)-低股利殖利率(最低10%)

HACLAC (營運資金率)

高營運資金率(最高10%)-低營運資金率(最低10%)

HEBITLEBIT 高EBIT(最高10%)- 低EBIT(最低10%)

UMD1 一個月動能因子

顯著的結果為
MKT,HML,CMA與QMJ
額外的因子都無顯著性

F(12,167) 32.01

R² 0.70

Adj. R² 0.68

| | Est. | S.E. | t val. | p | VIF |
|----------------------|-------|------|--------|------|------|
| (Intercept) | 0.41 | 0.24 | 1.70 | 0.09 | NA |
| MKT | 0.92 | 0.08 | 12.03 | 0.00 | 2.11 |
| SMB3 | 0.06 | 0.11 | 0.52 | 0.60 | 1.35 |
| HML | -0.42 | 0.13 | -3.09 | 0.00 | 2.56 |
| RMW | 0.23 | 0.22 | 1.05 | 0.30 | 2.31 |
| CMA | -0.56 | 0.20 | -2.84 | 0.01 | 1.67 |
| UMD | -0.03 | 0.07 | -0.36 | 0.72 | 2.34 |
| BAB | -0.04 | 0.10 | -0.45 | 0.65 | 1.25 |
| QMJ | -0.33 | 0.16 | -2.00 | 0.05 | 3.29 |
| HDPLDP | -0.03 | 0.10 | -0.36 | 0.72 | 1.80 |
| HACLAC | -0.10 | 0.15 | -0.66 | 0.51 | 1.24 |
| HEBITLEBIT | 0.09 | 0.13 | 0.70 | 0.48 | 2.50 |
| UMD1 | 0.10 | 0.08 | 1.37 | 0.17 | 1.27 |
| Standard errors: OLS | | | | | |

因子穩健性測試(牛/熊市)

利用以下殖利率區分牛市與熊市
並測試各因子在不同時期表現

$$\text{UST2y10y} = \text{UST10y YTM} - \text{UST2y YTM}$$

牛市: $\text{UST2y10y} > \text{UST2y10y 5年平均}$

熊市: $\text{UST2y10y} < \text{UST2y10y 5年平均}$

結果顯示 MKT,HML,CMA 在牛市時依然有顯著性

熊市時只有 MKT 與 CMA 時依然有顯著性

F(3,112) 107.15

R² 0.74

Adj. R² 0.73

F(2,61) 54.11

R² 0.64

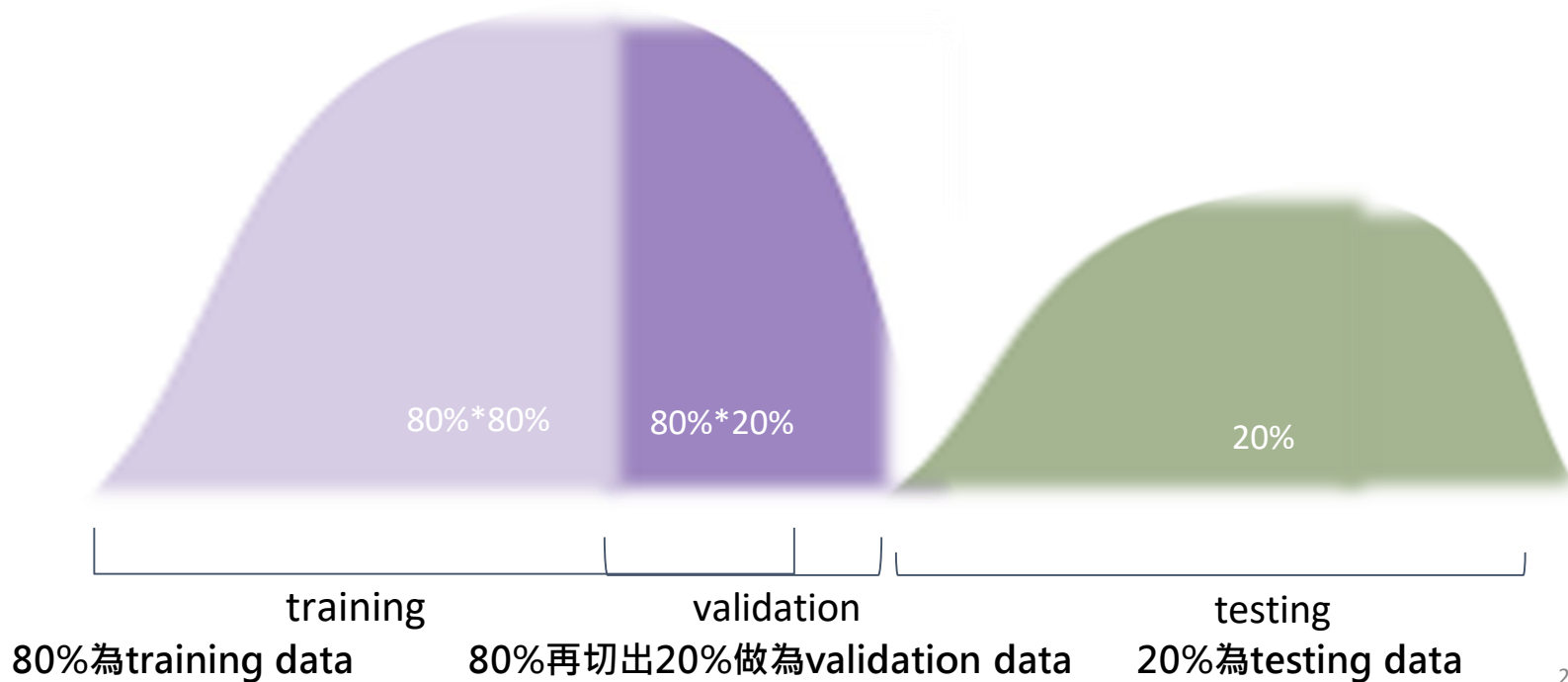
Adj. R² 0.63

| | Est. | S.E. | t val. | p | VIF |
|----------------------|-------|------|--------|------|------|
| (Intercept) | 0.17 | 0.19 | 0.88 | 0.38 | NA |
| MKT | 0.96 | 0.06 | 15.50 | 0.00 | 1.07 |
| HML | -0.32 | 0.10 | -3.12 | 0.00 | 1.45 |
| CMA | -0.42 | 0.17 | -2.56 | 0.01 | 1.53 |
| Standard errors: OLS | | | | | |

| | Est. | S.E. | t val. | p | VIF |
|----------------------|-------|------|--------|------|------|
| (Intercept) | 0.78 | 0.52 | 1.49 | 0.14 | NA |
| MKT | 0.95 | 0.09 | 9.99 | 0.00 | 1.00 |
| CMA | -1.17 | 0.36 | -3.23 | 0.00 | 1.00 |
| Standard errors: OLS | | | | | |

機器學習方式測試模型穩健性

除了以上因子模型(線性回歸)外，我們也利用機器學習的方式來測試此模型的表現及穩健度



建置模型

將MKT,HML,CMA三個因子放入四種機器學習模型檢測

val_score為validatin data跑模型後的R Square · val_e為validatin data跑模型後的MSE
test_score為testing data預測後的R Square · test_e為testing data預測後的MSE

SVR模型

| | c | val_score | val_e | test_score | test_e |
|---|--------|-----------|-----------|------------|-----------|
| 0 | 1000.0 | 0.758263 | 4.957593 | 0.836331 | 3.428702 |
| 1 | 1000.0 | 0.632661 | 5.821681 | 0.830201 | 4.514368 |
| 2 | 1000.0 | 0.099834 | 21.742462 | 0.847144 | 4.082156 |
| 3 | 1000.0 | 0.791601 | 6.923173 | 0.809425 | 4.145156 |
| 4 | 1000.0 | 0.365735 | 19.936032 | 0.857906 | 4.475574 |
| 5 | 1000.0 | 0.834000 | 4.623150 | 0.784122 | 5.424589 |
| 6 | 1000.0 | 0.730999 | 4.858144 | 0.562474 | 21.478693 |
| 7 | 1000.0 | 0.390006 | 21.862468 | 0.848637 | 3.995889 |
| 8 | 1000.0 | 0.784923 | 3.238209 | 0.796643 | 7.191452 |
| 9 | 1000.0 | 0.755878 | 4.158401 | 0.830779 | 4.116596 |

Lasso模型

| | alpha | val_score | val_e | test_score | test_e |
|---|----------|-----------|-----------|------------|----------|
| 0 | 0.003818 | 0.779119 | 3.235292 | 0.786671 | 6.529222 |
| 1 | 0.003916 | 0.857366 | 3.717379 | 0.816279 | 4.443472 |
| 2 | 0.004067 | 0.731846 | 4.389223 | 0.745899 | 4.963448 |
| 3 | 0.003839 | 0.715286 | 5.213152 | 0.808742 | 3.959026 |
| 4 | 0.004155 | 0.845207 | 4.325772 | 0.612077 | 6.193650 |
| 5 | 0.004405 | 0.620369 | 5.009145 | 0.858220 | 3.364066 |
| 6 | 0.003807 | 0.876584 | 3.999216 | 0.789498 | 5.499945 |
| 7 | 0.004483 | 0.686462 | 5.411398 | 0.599280 | 6.302248 |
| 8 | 0.003758 | 0.549110 | 23.395484 | 0.812549 | 5.206700 |
| 9 | 0.003982 | 0.647539 | 5.453763 | 0.724306 | 5.987409 |

建置模型

Ridge模型

| | alpha | val_score | val_e | test_score | test_e |
|---|-------|-----------|-----------|------------|-----------|
| 0 | 1.0 | 0.846740 | 4.020822 | 0.776784 | 4.225624 |
| 1 | 1.0 | 0.449264 | 21.471508 | 0.819362 | 4.609554 |
| 2 | 0.1 | 0.360010 | 22.887258 | 0.778610 | 4.800413 |
| 3 | 1.0 | 0.688632 | 6.609734 | 0.815154 | 5.322875 |
| 4 | 1.0 | 0.597143 | 23.748747 | 0.860658 | 2.920092 |
| 5 | 1.0 | 0.846579 | 3.193546 | 0.574766 | 6.293811 |
| 6 | 1.0 | 0.749613 | 4.297983 | 0.833270 | 4.192280 |
| 7 | 1.0 | 0.714046 | 5.051558 | 0.359472 | 21.974545 |
| 8 | 1.0 | 0.858234 | 4.514634 | 0.640881 | 7.706992 |
| 9 | 1.0 | 0.765997 | 4.440677 | 0.879909 | 4.452235 |

三種方法經權重後之模型

| | val R square | val e | prediction R square | prediction e |
|---|--------------|-----------|---------------------|--------------|
| 0 | 0.825401 | 4.820698 | 0.775087 | 4.298646 |
| 1 | 0.821160 | 3.258697 | 0.705081 | 5.586994 |
| 2 | 0.752171 | 5.838494 | 0.341074 | 19.836871 |
| 3 | 0.732937 | 3.945901 | 0.815987 | 3.773831 |
| 4 | 0.680507 | 6.830978 | 0.722139 | 4.546509 |
| 5 | 0.744692 | 3.676104 | 0.695234 | 5.814780 |
| 6 | 0.545189 | 20.903503 | 0.832131 | 3.667378 |
| 7 | 0.673349 | 4.819400 | 0.404918 | 20.943237 |
| 8 | 0.736165 | 3.936077 | 0.860527 | 5.282630 |
| 9 | 0.775516 | 4.347922 | 0.819837 | 5.554488 |

建置模型

平均比較

| | validation R square | validation_e | testing R square | testing_e | val R square std | test R square std |
|-------|---------------------|--------------|------------------|-----------|------------------|-------------------|
| svr | 0.691247 | 8.407391 | 0.677618 | 8.827650 | 0.181908 | 0.174074 |
| ridge | 0.705060 | 8.029673 | 0.676284 | 8.835902 | 0.181908 | 0.174074 |
| lasso | 0.705731 | 8.355257 | 0.667179 | 9.167487 | 0.181908 | 0.174074 |
| emb | 0.695490 | 8.489090 | 0.686410 | 8.390424 | 0.181244 | 0.178261 |

各演算法各跑500次的validation set及test set平均R square、MSE標準差
以及三種的算法的總平均(row:emb)
結果顯示測試組的配適度都與迴歸相近，足夠穩健



投資組合策略

股票交易特性與 Feature 介紹

由於並非所有投資人都為價值型投資，不能僅利用因子模型作為篩選
例如: 有些投資人是利用技術分析，或是投資特定議題相關之股票
為解決此問題，我們額外增加一層篩選，主要是利用股票的交易特性來進行模型訓練及篩選

01



Volatility

股票的波動率
計算方法為股價的標準差

liquidity

股票的流動性，計算方法為該月股票成交量/流通在外股數

03



Momentum

股票的動能，計算方法為前2個月股價收盤價/前12個月股價收盤價

52Weeks High

(自定義)
計算方法為 52周內最高價/52周內最低價

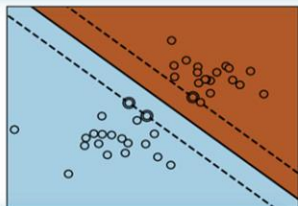
02



04



Features模型建置



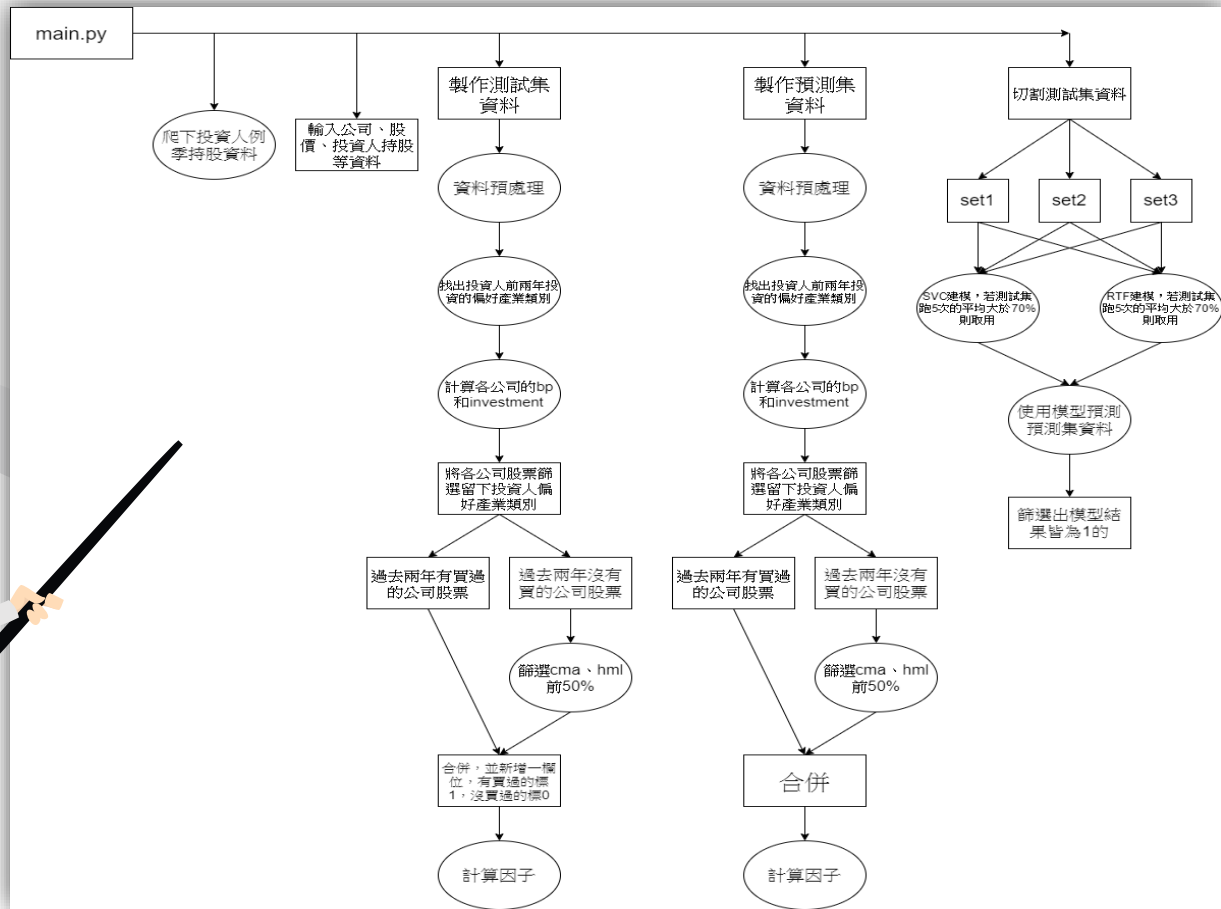
2維SVC示意圖

訓練時，放入上一期之features
並使在本期投資人有買之股票=1；
隨機從市場抽樣設為本期投資人未買的股票=0
最後得出數個模型

利用各模型作為我們下一期選股之模型，也就是在下一期買股票時，會利用此期features以及此模型挑選股票。

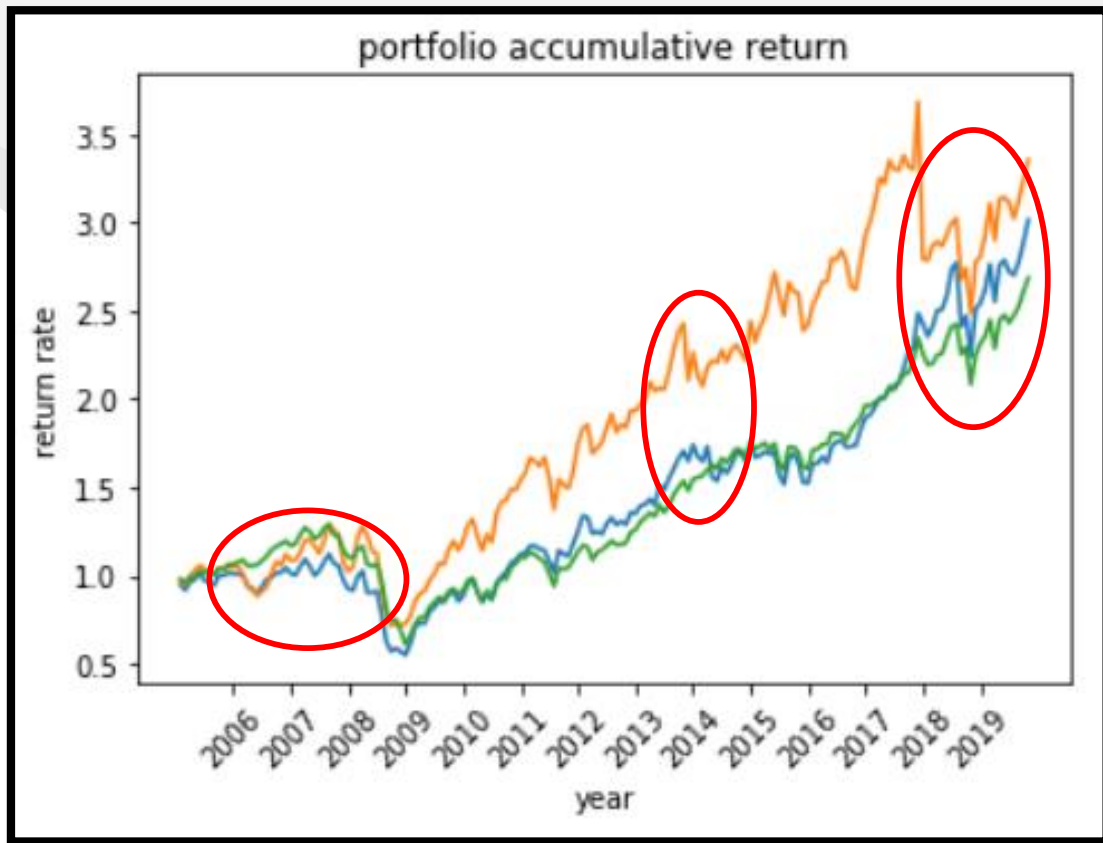
得出數組股票後，我們將其取交集，得知在下一期時我們會選哪些股票。投資組合採用Price weighted 方式組成

利用SVC分類模型及隨機森林訓練



Features訓練及篩選投組流程圖

模型結果



橘色 : Stephen Mandel Portfolio
(Price-weighted)

藍色 : Our Portfolio
(Price-weighted)

綠色 : S&P500

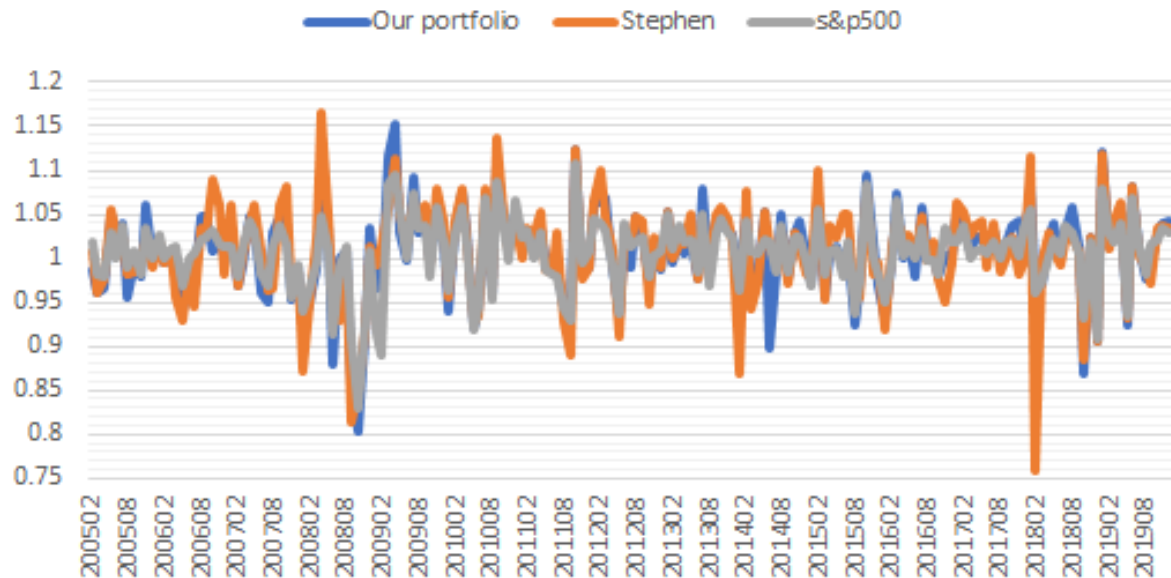
我們的投資組合累積報酬與Stephen Mandel的累積報酬相關性

0.927874

雖然我們的報酬看似與SP500接近。
但在 Stephen Mandel 的報酬與市場
走勢不同時，我們仍可以抓到其趨勢

我們推測可能原因為，在這些時間中
我們放入投資組合的股票數量較少，
因此會更接近投資人的表現，拉開與
大盤的走勢

每月報酬



從月報酬來看，Stephen的月報酬大多數都會高於我們的投組及s&p500數個百分點，長期複利效果造成雖然我們的相關性強，累積報酬卻會落後

我們預計再將加入不同Features，減少股票池數量，並嘗試不同權重

未來一周預計進度

- 由於目前我們的篩選投組中的股票數還是有點多，可能會造成結果與市場相似，因此我們首要的目標是在第三層多增加一些Features來篩選。
- 我們目前的方法是每股買一支，也就是price-weighted，績效相比其他權重法，通常表現較不好，各種研究皆已顯示價格加權法為較無效率之加權方式。因此我們接下來預計嘗試equal weighted的方式來進行投資
- Chat Bot更新(目前已有陽春版本)



結語

我們以Stephen Mandel作為我們的探討對象，分析其投資傾向並試圖複製該投資人投資組合與報酬，我們發現該投資人傾向「徹底了解企業之基本面能預判未來成長」的投資方式並買入優良的公司

我們利用13F報表了解Stephen Mandel投資的股票與每月報酬，接著透過三層濾網分析投資人投資風格並藉此建構出投資組合

第一步，我們分析Stephen Mandel習慣投資的產業類別，並將前幾大類別股設為我們的股票池

第二步，我們以傳統投資因子分析Stephen Mandel的投資風格，並從前面的股票池挑出風格相近的股

第三步，我們加入技術面指標分析Stephen Mandel的投資風格，並以機器學習模型探討投資人的投資風格，最後成功建立出投資組合。

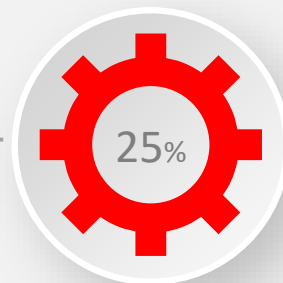
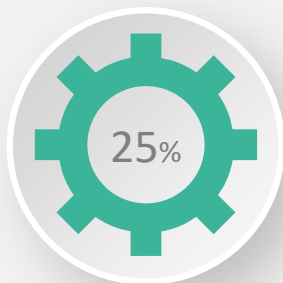
該投資組合從2004年至今的累計報酬為300%，與Stephen Mandel的累積報酬相關係數為0.927874，呈現高度相關。顯示出我們的投資組合篩選可以逼近Stephen Mandel的風格

許惟傑

簡報製作

架構規劃

尋找因子



盧煒中

資料爬蟲

資料清洗

機器學習

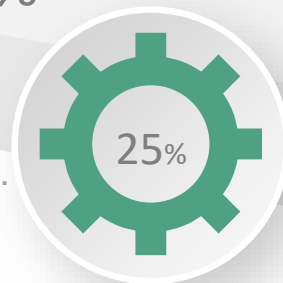
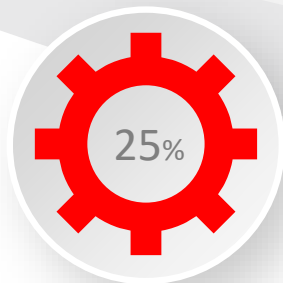
100%

林哲瑋

簡報製作

架構規劃

尋找因子



徐慕華

簡報製作

架構規劃

尋找因子

THANK YOU

Machine
Learning

Celebrity

FACTO
R

CTBC

