

玉山第二組 專題匯報

指導業師：玉山證券 詹益安

指導老師：台大 張智星教授

東吳 蔡芸琇老師

台大材料二 B08507031 黃子騰

東吳巨資二A 08170110卓大一

東吳巨資二A 08170111尚旻儀

東吳巨資二A 08170125雷沅翰

分工狀況

黃子騰

違約交割預測模型建構

卓大一

尋找股票市場可能的風險

尚旻儀

研究API串接

雷沅翰

aws串接

挑戰

- 數據量龐大且特徵多，而許多資料有缺值需進行處理和整理
- beta值為系統性風險評估工具，對於非系統性風險無法衡量
- 資料雖多龐大，但違約交割標記客戶**數量少且佔比低**(僅96筆約0.09%)
- 針對違約交割行為出現較無明顯訊號，預測模型建構不易
- 因模型用於預測具違約交割風險之顧客，因此Recall需努力升高

產品功能

違約交割預測模型

協助公司及使用者瞭解客戶是否有違約交割風險

購買標的風險評估模型

協助使用者理解購買標的風險

違約交割預測模型 - 建構流程



違約交割預測模型 - 資料前處理

使用Pandas進行大部分數據整理和統計，也用於產出新特徵

1. 個股資料

- 對應日期和個股代碼，將每日beta21值加入
- 針對類股進行每日beta值平均計算（因無市值資料，無進行加權）

2. 客戶資料

- 刪除交易次數小於10次之顧客資料
- 新增購買股票種類統計（上市櫃、興櫃），並加以 normalized
- 新增交易方式統計（一般、當沖、信用交易），並加以 normalized
- 新增客戶購買標的beta21值平均，以及標的與類股beta21差值平均

違約交割預測模型 - 特徵選擇

1. 自變數特徵選擇

使用sklearn 中 `train_test_split`和`cross_val_score`進行資料切割和特徵選取, 最終選取四個特徵得到初步較佳預測力

- `beta_2l_mean`: 客戶購買標的`beta2l`值平均
- `Day`: 經`normalized`之當沖次數
- `BUY_COUNT`: 經`normalized`之購買次數
- `BETA`: 客戶購買標的與類股`beta2l`差值平均

2. 樣本資料平衡

因違約交割標記客戶數量少且比例低, 使用`Imbalanced-learn`中`SMOTE`套件平衡訓練樣本, 調整至違約交割至1:0.8

違約交割預測模型 - 模型建立和調整

1. 模型選擇

嘗試使用數個模型進行分類分析，模型選擇上主要參考以下因素：

- 訓練時間
- 避免過擬合
- 易於理解 and 解釋

參考以上數點後最終使用**決策樹**，且更進一步使用集成學習以提升預測準確率，包含：

- Random Forest
- XGBoost

2. 超參數調整

- `n_estimators`
- `max_depth`
- `min_child_weight`

違約交割預測模型 - 模型解釋和串接UI

1. 模型解釋

由於此模型用於預測違約交割客戶的出現，因此Recall值十分重要，且出現False Positive也可視為用於預備違約行為出現

- Recall: 45.4%

約可抓取一半違約標記客戶

- Accuracy: 99.5%

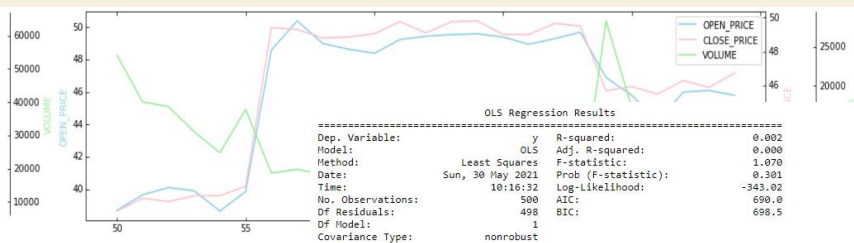
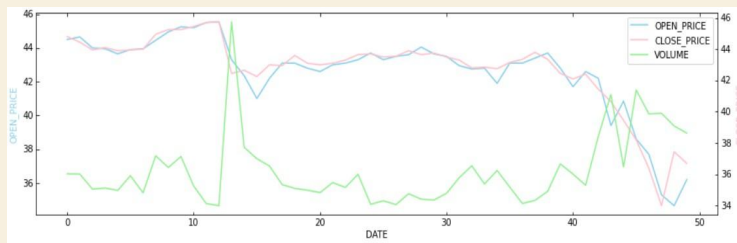
因無違約交割風險的客戶佔比極高，因此容易取的高辨識率

2. 串接UI

使用者輸入其客戶代號後即可獲得的交易記錄統計及beta值偏好程度

並會抓取該客戶資料帶入模型預測是否有違約交割風險

階段成果-股票指標(一)



OLS Regression Results

Dep. Variable:	y	R-squared:	0.002
Model:	OLS	Adj. R-squared:	0.000
Method:	Least Squares	F-statistic:	1.070
Date:	Sun, 30 May 2021	Prob (F-statistic):	0.301
Time:	10:16:32	Log-Likelihood:	-343.02
No. Observations:	500	AIC:	690.0
Df Residuals:	498	BIC:	698.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	0.0235	0.046	0.507	0.612	-0.068	0.115
x1	1.182e-07	1.14e-07	1.034	0.301	-1.06e-07	3.43e-07

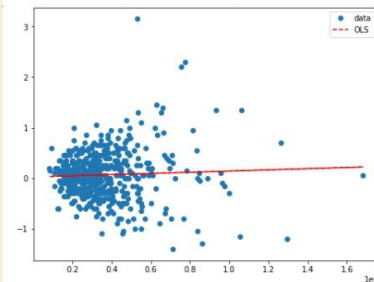
Omnibus: 94.039 Durbin-Watson: 1.879
Prob(Omnibus): 0.000 Jarque-Bera (JB): 474.608
Skew: 0.712 Prob(JB): 8.71e-104
Kurtosis: 7.555 Cond. No. 8.74e+05

• 震倉

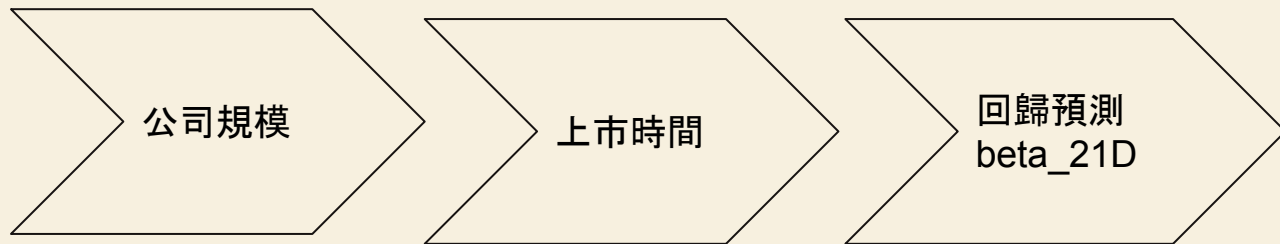
大多時候我們都關注股價，但有時關注交易量反而更為精準，這時可能代表資金主力可能要退場了，之後股價反而

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 8.74e+05. This might indicate that there are strong multicollinearity or other numerical problems.

<matplotlib.legend.Legend at 0x1c8fad1e0>



階段成果(二)-修正方向



預想方案：

OLS (Original Least Square): 最小化誤差的平方和, 尋找最佳函數

ARIMA(Autoregressive Integrated Moving Average model)

Ridge regression(嶺回歸)

規模係數 & 時間係數

(I) 公司規模係數(par):

小規模公司穩定度相對於大公司低

(II) 時間係數(date_rank):

創立時間愈久, 內部流程相對完善

```
a = input(str('你買了哪支股票: '))

print('你買了:', a)

fr_1 = (df['STOCK_NO']==a)

date_rank = df[fr_1]['DATE_RANK'].count()

if df[fr_1]['CAPITAL_TYPE'][-1:].item() == 'H':
    par = 1
elif df[fr_1]['CAPITAL_TYPE'][-1:].item() == 'M':
    par = 1.05
else: par = 1.1

i = math.log(date_rank,561)
i1 = 1/i
```

嶺回歸

(III)預測顧客可能接受之beta值

```
X = df[fr_1]['CLOSE_PRICE'].to_numpy()
y = df[fr_1]['beta_21D'].to_numpy()

clf_ridge=linear_model.Ridge(alpha=0.5,tol = 0.1, solver='auto',normalize=True)
clf_ridge.fit(X.reshape(-1,1), y.reshape(-1,1))

print('最近五天的收盤價為: ', df[fr_1]['CLOSE_PRICE'].tail().tolist(), '平均價格為: ', df[fr_1]['CLOSE_PRICE'].tail().mean())
b = float(input('what price u buy: '))
b = np.array([b])
result = clf_ridge.predict(b.reshape(-1,1))

print(df[fr_1]['Industry'][-1:].tolist(), '該類股最近一天的總成交量為:', df[fr_1]['AMOUNT_y'][-1:].tolist())

print('風險係數為: ', par * i1 * result)
```

API研究-telegram

1. 完成資料上的整併

2. 查詢的功能

3. 大概功能

/start

/help

/about

4. 影片連結:

AWS串接

訊息詳細資訊

主題 ARN
arn:aws:sns:us-east-1:983848544876:Telegram

主旨 - [選用](#)

長度上限為 100 個可列印的 ASCII 字元

存留時間 (TTL) - [選用](#)
此設定僅適用於行動應用程式端點。推播通知服務將訊息交付至端點所需的秒數。 [資訊](#)

訊息內文

訊息結構

☒ 所有交付協定承載相同。
無論其交付協定如何，都將相同的承載傳送到訂閱該主題的端點。

☐ 為每個交付協定自訂承載。
根據其交付協定，向訂閱該主題的端點傳送不同的承載。

要傳送到端點的訊息內文

1 Hello

以聊天機器人跟AWS的SNS做結合，可以達到主動推播的功能，在Lambda裡面設定ID跟Token，程式碼會在相對應的地方抓資料，就可以傳訊息到對應的聊天室。