

量化投資實務 期末報告

臺科大財金所

碩士生：李翔緯

許絹書. (2017). 授信風險模型之評估-以某租賃公司為例[碩士論文, 國立臺北大學會計學系]. 國家圖書館.
<https://ndltd.ncl.edu.tw/cgi-bin/gs32/gswweb.cgi/login?o=dnclcdr&s=id=%22105NTPU0385038%22.&searchmode=basic>

目錄

- 研究目的
- 研究對象
- 研究變數
- 研究方法
- 研究結果
- 研究反思

研究目的

2022年台灣半導體業產值

項目	內容
台灣半導體業產值	4.89兆元
台灣半導體產業附加價值	2.98兆元
占GDP比重	13.1%
半導體從業人數	32.7萬
每人平均創造產值	900萬元
台積電占台灣GDP比重	7.94%

資料來源：台灣半導體產業協會

蘇嘉維 / 製表

- 期中報告的研究是探討個案公司所承作融資租賃及分期付款買賣契約等業務，並以該授信條件為基礎，利用 Logistic Regression 訂定出有效的授信風險之模型。
- 故以此為基礎採用更為嚴謹的方式來建立本國電子產業的違約授信模型

研究對象：TSE 24,25,26,27,30

- 囊括大部分電子產業

21:生物科技業
22:通訊網路業
23:油電燃氣
24:半導體業
25:電腦及週邊設備業
26:光電業
27:通訊網路業
28:電子零組件
29:電子通路業
30:資訊服務業
31:其他電子業
32:文化創意業
33:農業科技

- 用上下資本額10%進行配對
- 若配對不成擴大到20%進行配對

危機發生日	證券代碼	簡稱	TSE	產業別	實收資本額	0.9	1.1	0.8	1.2
20080131	8094	卓立	27	cap_1	318280030	286452027	350108033	254624024	381936036
20080425	3096	碩貝	24	cap_2	350525320	315472788	385577852	280420256	420630384
20080430	5414	磐英	25	cap_3	685839170	617255253	754423087	548671336	823007004
20090317	8150	南茂	24	cap_1	7272401260	6545161134	7999641386	5817921008	8726881512
20091218	4919	新唐	24	cap_2	4197652680	3777887412	4617417948	3358122144	5037183216
20100831	6149	禾鴻	30	cap_1	318796000	286916400	350675600	255036800	382555200
20100913	3661	世芯	24	cap_2	798880200	718992180	878768220	639104160	958656240
20101220	4994	傳奇	30	cap_3	663834000	597450600	730217400	531067200	796600800
20111220	3658	漢微	24	cap_1	710000000	639000000	781000000	568000000	852000000
20120831	4992	宇通	26	cap_1	2100500000	1890450000	2310550000	1680400000	2520600000

• 違約：52

• 非違約：542

研究變數：量化與類別與標準化

- 財務資料

- 自行計算一些量化比率

- 營運現金流量佔本期產生現金流量
- 應付帳款及票據佔流動負債
- 普通股股本佔股東權益總額
- 營運指數
- 營運資金周轉率
-

- ESG資料

- 環境保護

- 受限於資料期間

- 社會責任

- 高管人數大於3人
- 可以再加入

- 公司治理

- 3年內CPA是否換人
- 3年內是否更換董事長
- 3年內總經理有無異動

研究方法：單因子分析(Wald test)

Variable		P-value	# P-value_numeric
Missing:	0 (0%)	Missing:	0 (0%)
Distinct:	79 (100%)	Distinct:	38 (48%)
79 Distinct values		<0.001: 0.0017: 0.5784: Other:	49% 3% 3% 46%
		Min 0.001 Max 0.9741	
78	3年內內部稽核是否異動	<0.001	0.001
22	已實現銷售毛利率	<0.001	0.001
23	營業利益率	<0.001	0.001
24	稅前淨利率	<0.001	0.001
25	稅後淨利率	<0.001	0.001
26	業外收支/營收	<0.001	0.001
27	常續利益率 - 稅後	<0.001	0.001
28	貝里比率	<0.001	0.001
29	營業資產報酬率	<0.001	0.001
30	營業費用率	<0.001	0.001
31	用人費用率	<0.001	0.001
33	現金流量比率	<0.001	0.001
34	常續性EPS	<0.001	0.001
36	每股營業額	<0.001	0.001
49	淨值成長率	0.0010	0.001
46	經常淨利成長率	<0.001	0.001
77	3年內發言人有無異動	<0.001	0.001
45	稅後淨利成長率	<0.001	0.001
44	稅前淨利成長率	<0.001	0.001
20	ROE(B) - 常續利益	<0.001	0.001

- 找每個變數的P-value
- 將P-value轉數值以排序和比較
- 使用函數將P-value<0.05的保留

原始變數數量: 79

顯著變數數量: 52

移除的變數 (P值 ≥ 0.05): ['季底應收帳款/營收TSE'],

保留的變數 (P值 < 0.05): ['3年內內部稽核是否異動'],

研究方法：相關性分析+VIF?



• 相關係數矩陣

- 將相關係數絕對值大於0.5的變數使用兩層的for迴圈刪除

=== 相關係數分析結果 ===

原始變數數量: 52

高相關變數數量: 26

保留變數數量: 26

移除的變數: ['ROA(B)稅後息前折舊前',

保留的變數: ['3年內內部稽核是否異動',

研究方法：逐步回歸

```
# 逐步回歸限制版(10個變數 7個財務變數 3個類別變數)
# 讀取財務變數和類別變數
fin_vars = pd.read_csv('fin_x2.csv').columns.tolist() # 財務變數列表
dummy_vars = pd.read_csv('dum_x.csv').columns.tolist() # 類別變數列表

def stepwise_selection(x, y, financial_vars, dummy_vars, initial_list=[],
                      threshold_in=0.5, threshold_out=0.5,
                      max_total=12, max_financial=7, max_dummy=2, verbose=True):

    included = list(initial_list)
    removed = set()

    while True: ...

    # 如果超過限制，分別處理財務和類別變數
    model = sm.Logit(y, sm.add_constant(pd.DataFrame(X[included]))).fit(dispen=0)
    pvalues = model.pvalues.iloc[1:]
```

- 使用while迴圈將變數限制在10個
- 再剔除不顯著變數

```
=== 逐步迴歸分析結果 ===
原始變數數量: 26
原始財務變數數量: 20
原始類別變數數量: 6
保留變數數量: 9
保留財務變數數量: 7
保留類別變數數量: 2
```


研究結果：Logit result

變數對照表：

x1: 總負債/總淨值

x2: 應付帳款及票據佔流動負債

x3: 經常淨利成長率

x4: 淨值成長率

x5: 現金及約當現金佔流動資產

x6: 業外收支/營收

x7: 普通股股本佔股東權益總額

x8: 3年內財務主管有無異動

x9: 3年內董事長有無異動

- 變動一單位=>變動勝算比

```

=====
Dep. Variable:                Y      No. Observations:                482
Model:                        Logit   Df Residuals:                    472
Method:                       MLE    Df Model:                        9
Date:                         Wed, 11 Dec 2024  Pseudo R-squ.:                0.5121
Time:                         22:09:16    Log-Likelihood:                 -69.580
converged:                     True      LL-Null:                        -142.61
Covariance Type:              nonrobust  LLR p-value:                    5.793e-27
=====

```

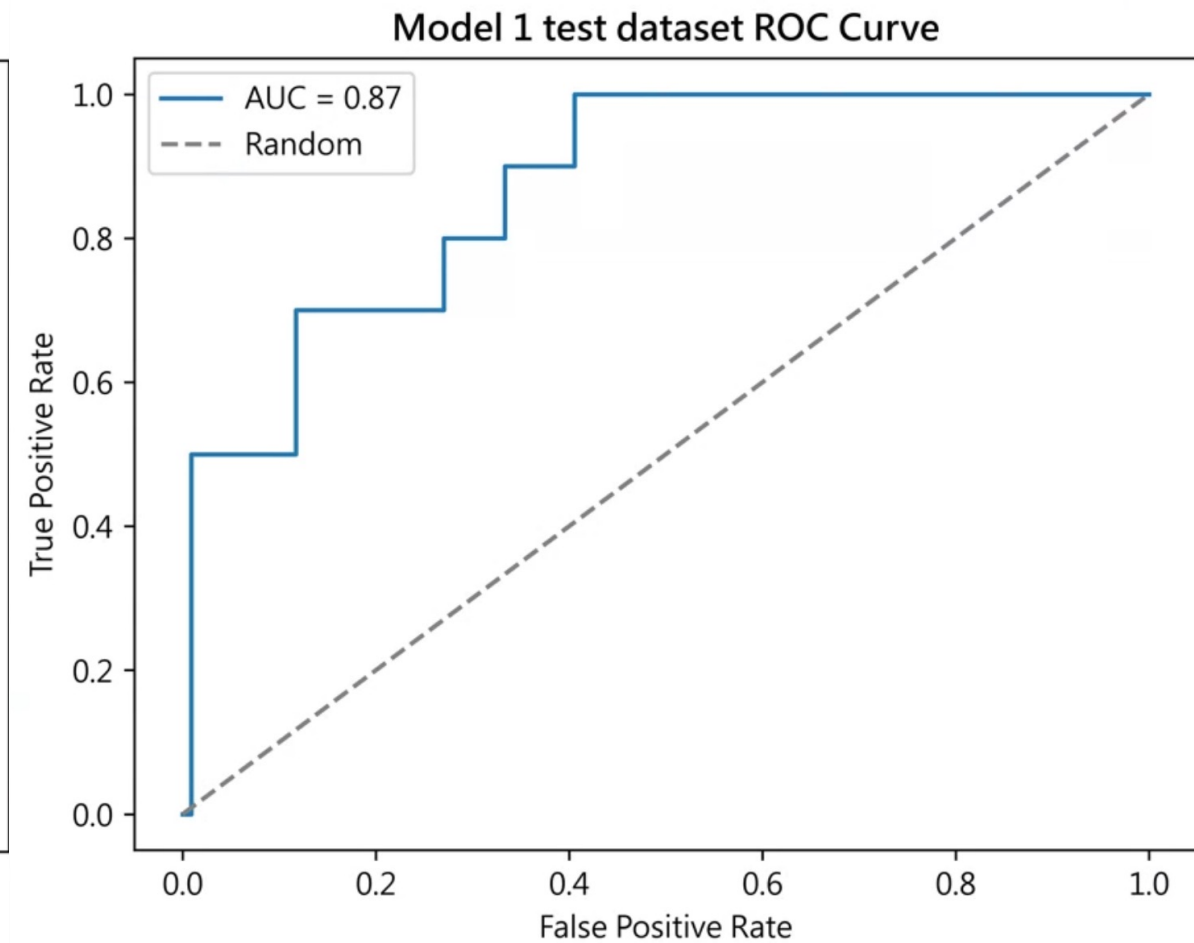
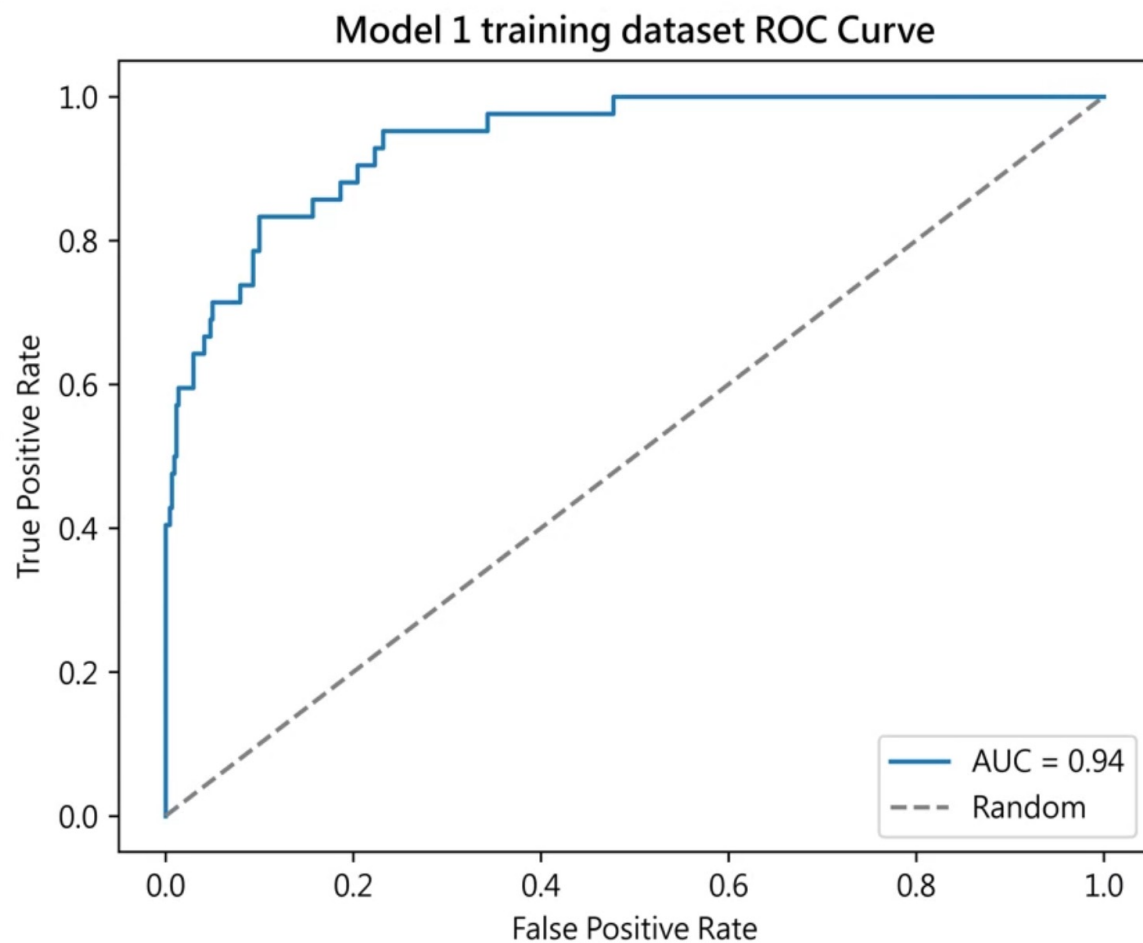
	coef	std err	z	P> z	[0.025	0.975]
const	-4.4375	0.528	-8.399	0.000	-5.473	-3.402
x1	4.1415	0.840	4.929	0.000	2.495	5.788
x2	-1.3362	0.337	-3.963	0.000	-1.997	-0.675
x3	-2.0043	0.518	-3.869	0.000	-3.020	-0.989
x4	4.0926	0.848	4.828	0.000	2.431	5.754
x5	-0.8395	0.281	-2.990	0.003	-1.390	-0.289
x6	-1.7099	0.560	-3.052	0.002	-2.808	-0.612
x7	-2.2382	0.738	-3.033	0.002	-3.685	-0.792
x8	1.4665	0.487	3.010	0.003	0.512	2.421
x9	1.4568	0.595	2.447	0.014	0.290	2.624

```

=====

```

研究結果：ROC (0.1)



研究結果：準確率0.9 & type1 error最低

Cut-off: 0.10
Confusion Matrix:
[[391 49]
[7 35]]
Type I Error (偽陰性率): 0.1667
Type II Error (偽陽性率): 0.1114
Accuracy (準確率): 0.8838
Precision (精確率): 0.4167
Recall (召回率): 0.8333
Total Errors (總錯誤數): 56

Cut-off: 0.15
Confusion Matrix:
[[403 37]
[11 31]]
Type I Error (偽陰性率): 0.2619
Type II Error (偽陽性率): 0.0841
Accuracy (準確率): 0.9004
Precision (精確率): 0.4559
Recall (召回率): 0.7381
Total Errors (總錯誤數): 48

Cut-off: 0.20
Confusion Matrix:
[[418 22]
[12 30]]
Type I Error (偽陰性率): 0.2857
Type II Error (偽陽性率): 0.0500
Accuracy (準確率): 0.9295
Precision (精確率): 0.5769
Recall (召回率): 0.7143
Total Errors (總錯誤數): 34

Cut-off: 0.30
Confusion Matrix:
[[424 16]
[15 27]]
Type I Error (偽陰性率): 0.3571
Type II Error (偽陽性率): 0.0364
Accuracy (準確率): 0.9357
Precision (精確率): 0.6279
Recall (召回率): 0.6429
Total Errors (總錯誤數): 31

Cut-off: 0.40
Confusion Matrix:
[[430 10]
[17 25]]
Type I Error (偽陰性率): 0.4048
Type II Error (偽陽性率): 0.0227
Accuracy (準確率): 0.9440
Precision (精確率): 0.7143
Recall (召回率): 0.5952
Total Errors (總錯誤數): 27

Cut-off: 0.50
Confusion Matrix:
[[434 6]
[18 24]]
Type I Error (偽陰性率): 0.4286
Type II Error (偽陽性率): 0.0136
Accuracy (準確率): 0.9502
Precision (精確率): 0.8000
Recall (召回率): 0.5714
Total Errors (總錯誤數): 24

Cut-off: 0.60
Confusion Matrix:
[[437 3]
[23 19]]
Type I Error (偽陰性率): 0.5476
Type II Error (偽陽性率): 0.0068
Accuracy (準確率): 0.9461
Precision (精確率): 0.8636
Recall (召回率): 0.4524
Total Errors (總錯誤數): 26

Cut-off: 0.70
Confusion Matrix:
[[440 0]
[25 17]]
Type I Error (偽陰性率): 0.5952
Type II Error (偽陽性率): 0.0000
Accuracy (準確率): 0.9481
Precision (精確率): 1.0000
Recall (召回率): 0.4048
Total Errors (總錯誤數): 25

研究結果：混淆矩陣

1 training datatest Confusion Matrix
Accuracy: 0.9004
Type I Error: 0.2619
Type II Error: 0.0841



1 test datatest Confusion Matrix
Accuracy: 0.8595
Type I Error: 0.4000
Type II Error: 0.1171



研究反思：E 構面的變數問題

證券代碼		年	總筆數	總金額	營建工程	移動污染	水污染次	一般廢棄	事業廢棄
779	國票證券	2010	1	1200	0	0	0	1	0
1101	台泥	2010	13	436600	0	3	2	2	5
1102	亞泥	2010	10	406000	0	4	2	0	1
1104	環泥	2010	11	786000	0	2	5	0	1
1108	幸福	2010	7	1310000	0	0	2	0	0
1109	信大	2010	1	10000	0	1	0	0	0
1111	欣欣水泥	2010	3	210000	0	0	1	0	0
1201	味全	2010	1	3000	0	1	0	0	0
1203	味王	2010	1	1200	0	0	0	1	0
1210	大成	2010	9	1312000	4	0	1	0	2
1215	卜蜂	2010	1	5000	0	1	0	0	0
1216	統一	2010	18	569000	0	10	1	0	2
1217	愛之味	2010	2	11000	0	1	0	0	1
1218	泰山	2010	3	15000	0	3	0	0	0
1219	福壽	2010	2	4200	0	1	0	1	0
1220	台榮	2010	2	175000	0	1	1	0	0
1227	佳格	2010	2	7200	0	0	0	1	1
1231	聯華食	2010	2	103000	0	1	0	0	0

- 數據集從2010年開始
- 無法配對內模型2007-2009年的數據
- 解決方法：
- 直接不使用
- 用0去填補
- 用機器學習的方式將模型內其他特徵抓取去預測2010年以前的資料