數據分析實戰(B)：Risk premium 風險溢酬迴歸預測

第七組

**1082414張恆達**

**1082404林靜慧**

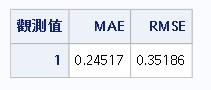
**1082417林宥萱**

**1082501賴柏薰**

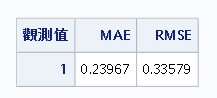
1. **最終模型最佳估計係數及回歸結果表**



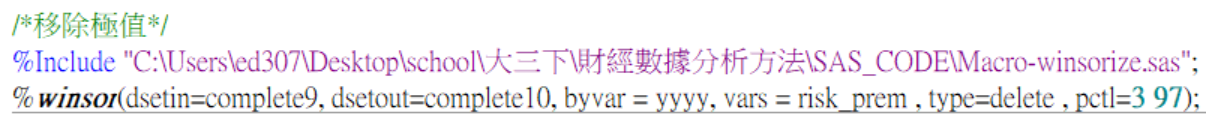
1. **最終模型 2020年預測準確度**



1. **最終模型 2021年預測準確度**

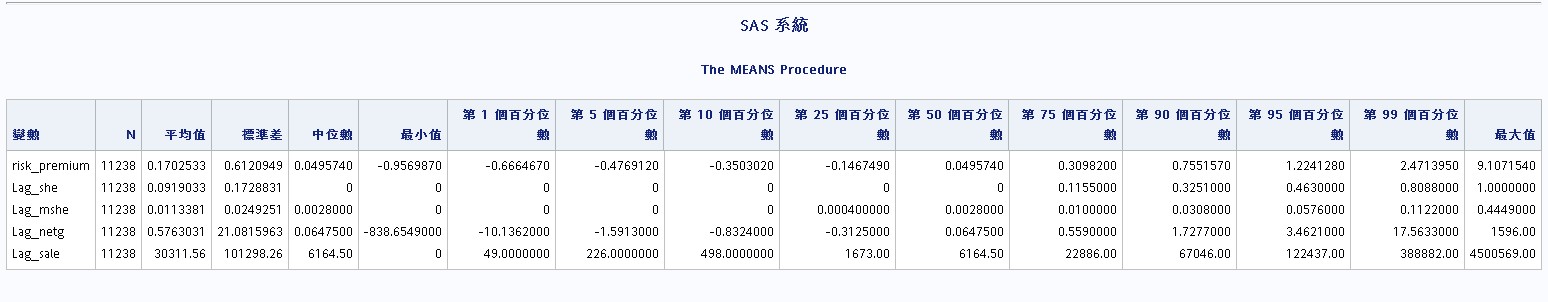


1. **最終模型之各變數定義**
2. **經理人持股％ ：**經理人持股數 / 總股數 \* 100%。公司高層，選取理由同董監持股％。
3. **稅後淨利成長率：**（税後淨利 / 去年同季稅後淨利）-1 \* 100%。可看出公司稅後淨利成長的情況。公司的營業成長高，未必代表稅後淨利成長率也高。因此稅後淨利成長率更能看出企業真實經營能力和效率。
4. **前一年該股票的非系統風險(idiosyncratic):**當年度 12 個月，月報酬去計算 的"個股 CAPM 模型得到的的非系統風險(systematic risk, beta)。Ri=b0+beta\*Rm+e。非系統風險即為此殘差的標準差。
5. **董監質押股％ ：**董監持股質押張數/ 總股數x 100%。如果董監事質押比例過高，其信用風險也偏高，表示董監事自身利益，與股東、公司整體利益脫節。
6. **極端值處理**

篩除前後各3％的極值，最終取出3%~97%之間的數值。

1. **敘述統計表**

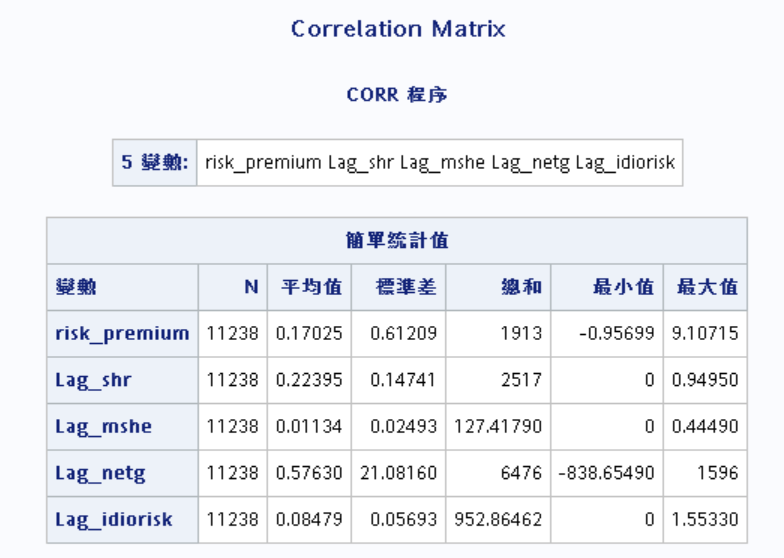
(1) 『最終模型』各變數(去除極端值前)



(2) 『最終模型』各變數(去除極端值後)



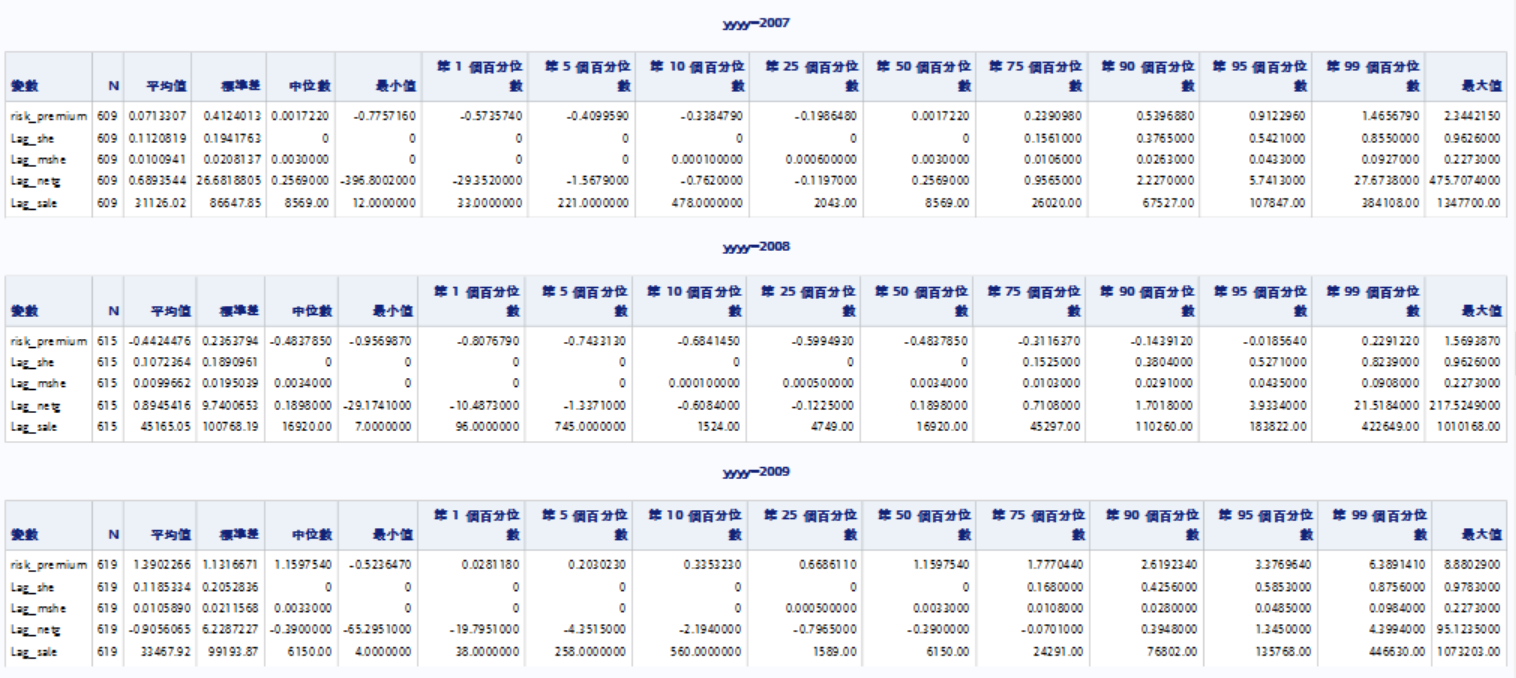
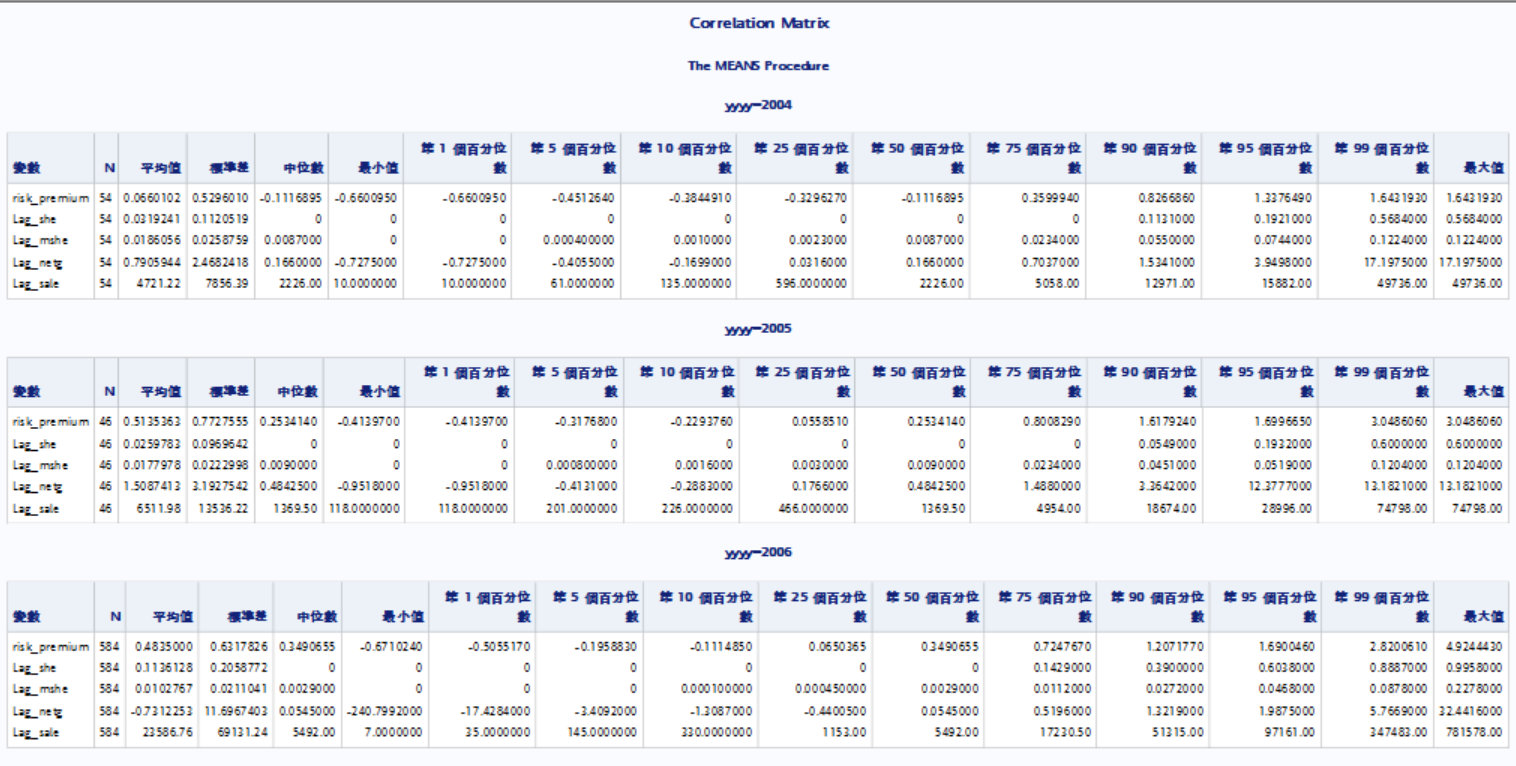
(3) 『最終模型』Pearson 相關係數各變數(去除極端值後)

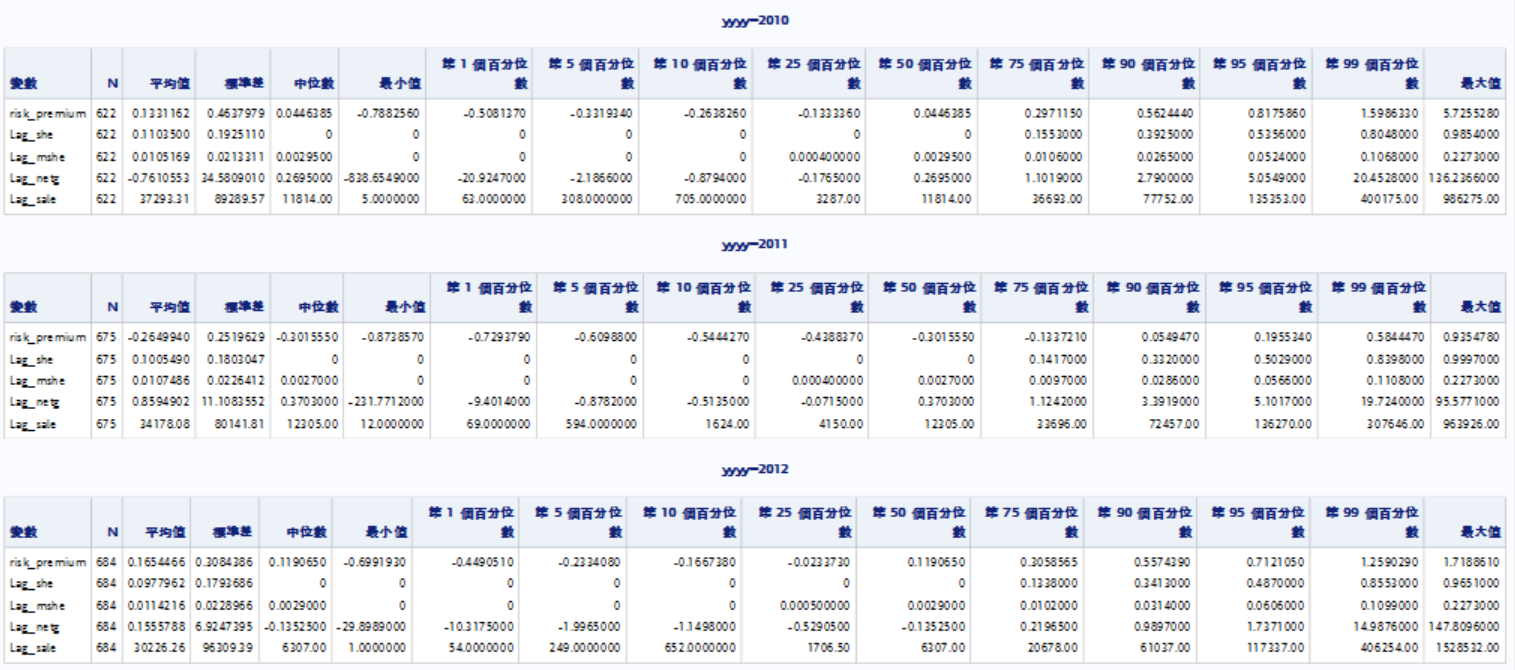


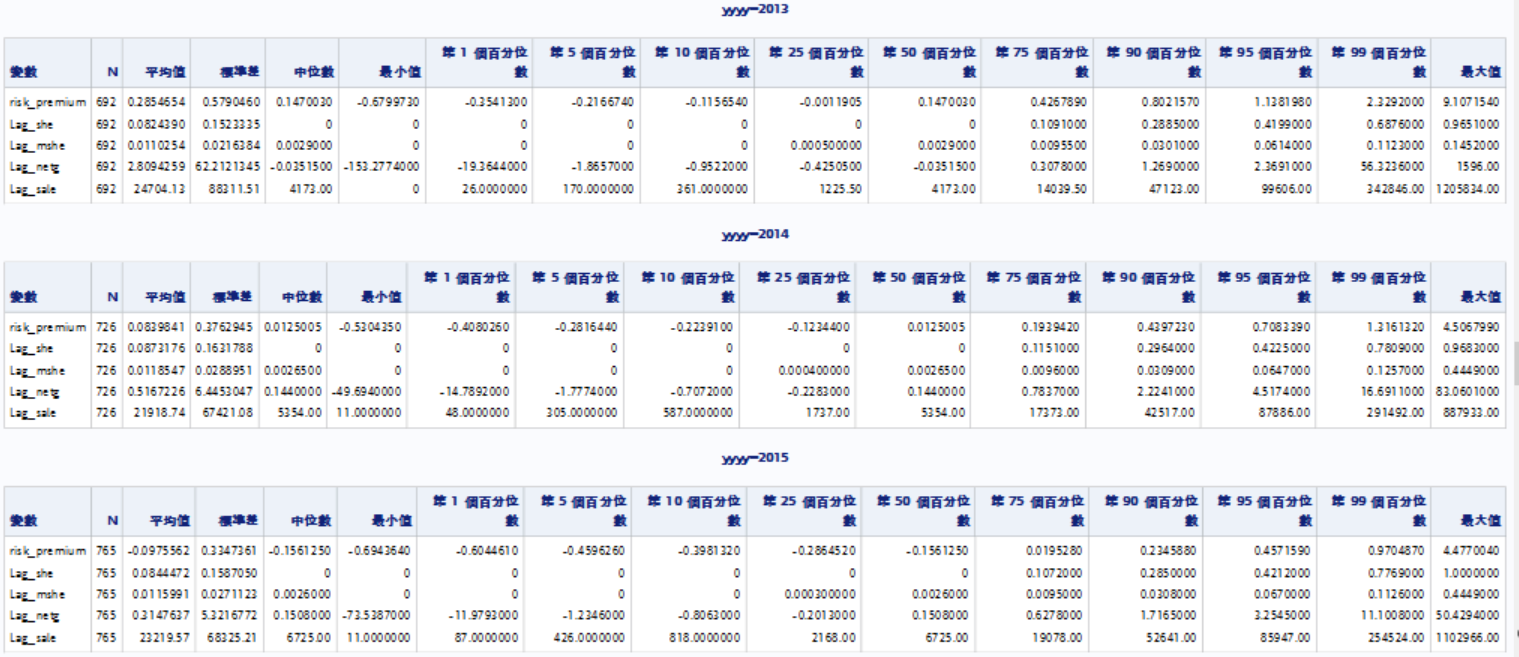


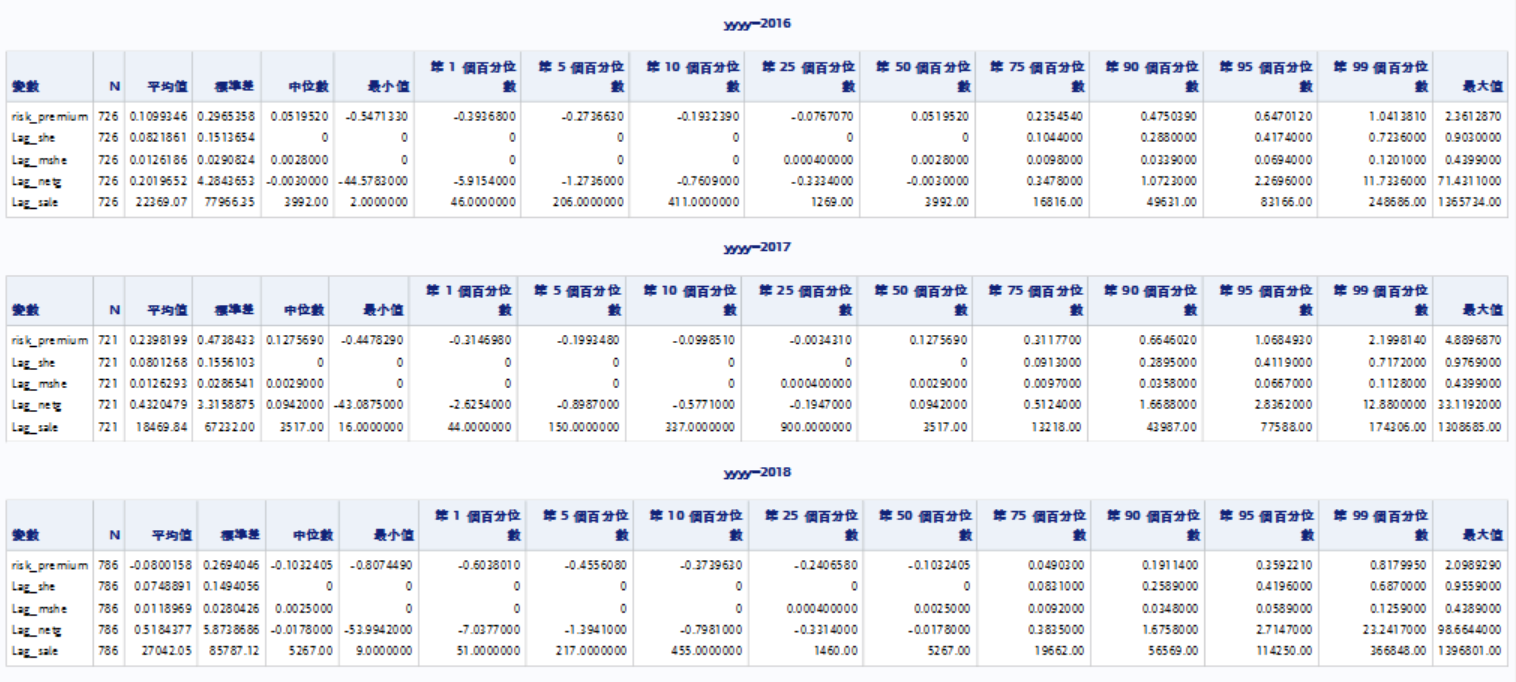
1. **排除各迴歸變數敘述統計表**

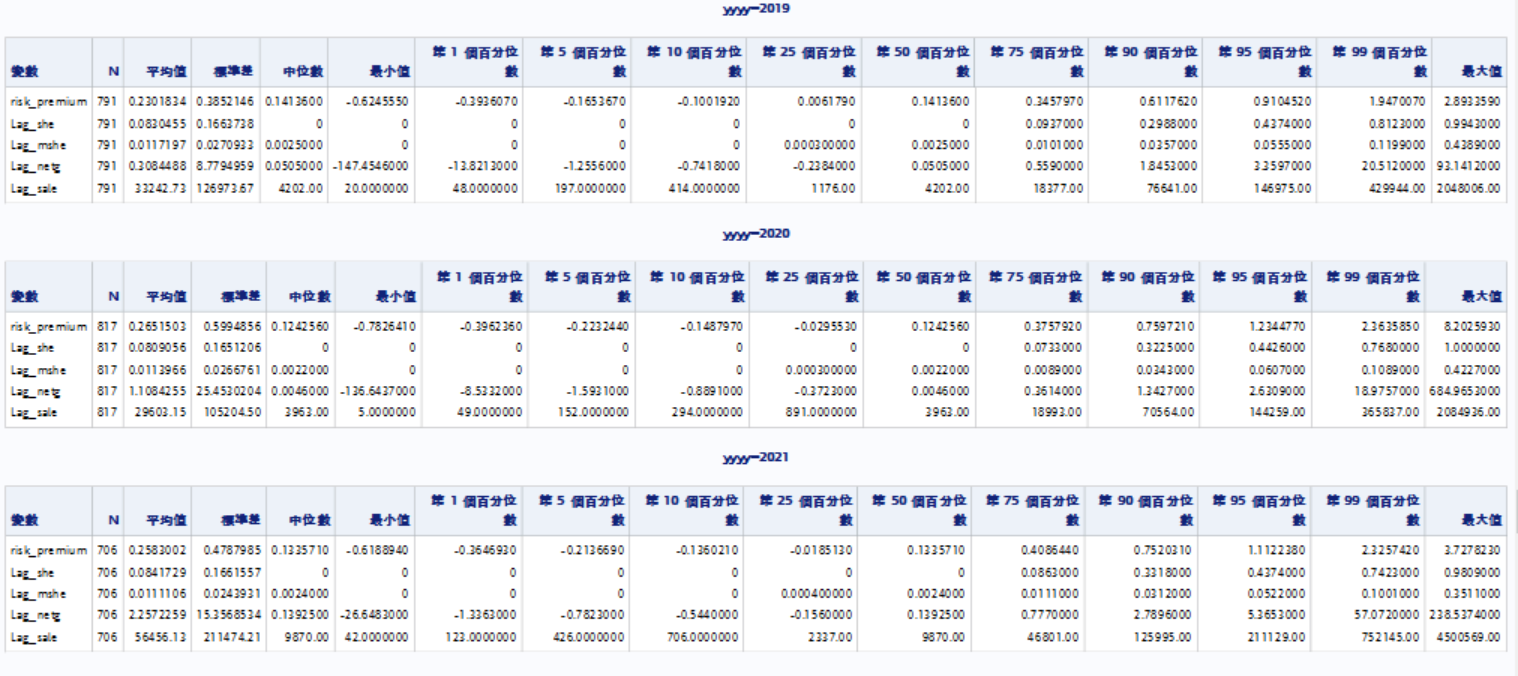
下表為 2004~2021 年，最終模型的迴歸內，排除各迴歸變數(應變數與自變數)的 missing value 後，應變數在處理極端值後，每年的敘述統計表，但由於2002-2003TEJ資料嚴重缺失，許多公司都為空值，所以只考慮2004~2021，故無敘述統計表。







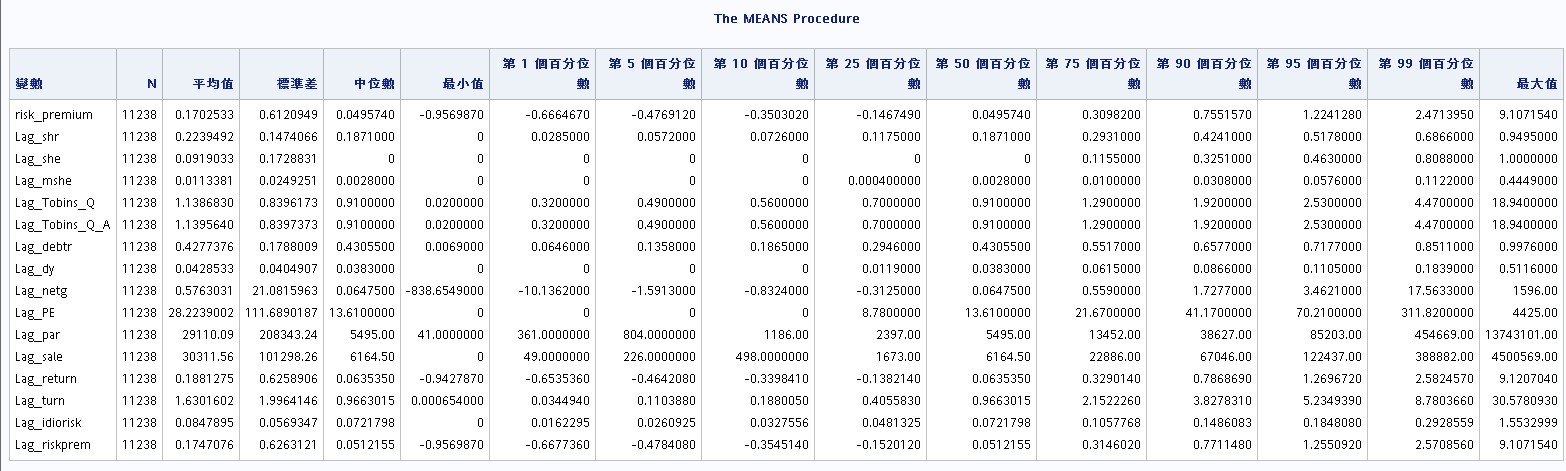




1. **最終模型中預期變數的回歸**



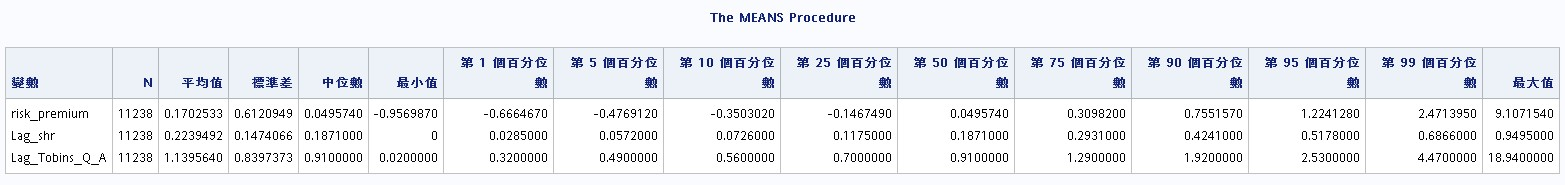
1. **董監質押比率 (shr):**董監質押比率越高，股價會硬撐上揚。因為一旦股價下跌，會讓抵押品的價值下降，所以董監事會想盡辦法避免股價下跌，質押比例大幅提高，股價可能會有很強的支撐，因此預計董監質押比率與股價的相關係數為正，而程式驗證結果也符合預期，因此保留。
2. **經理人持股 (mshe)：**經理人為公司內部高層，當經理人持股比率越高，代表對公司具有信心力，因此預計經理人持股與股價為相關係數為正，而程式驗證結果也符合預期，因此保留。
3. **淨成長率 (netg):**公司當年淨收益為正，代表公司穩定成長，內部經營無問題，因此同樣預計與股價呈正相關，但程式驗證結果不符合預期，我們還是選擇保留，因為股價呈現是未來資訊，現階段成長率為正並不會反映在現階段的股價上，另外保留原因還有淨成長率會是投資人選擇投資標的的考量之一。
4. **非系統風險 (idiorisk):** 非系統性風險又為可規避的風險，讓公司股價更穩定，因此跟市場呈現正相關。
5. **篩選嘗試模型變數之理由**
6. **董監持股％ :**董監事持股股數/ 總股數\* 100%。董監事為公司內部人，是最了解一間公司營運狀況之人。董監事持股的增加，會加強市場對於公司的信心，自然股價較不易下跌，故董監事持股比例為判斷股票是否值得投資的一項重要指標。
7. **股利殖利率(dy):** 現金股利 ÷ 股價。股價和現金股利兩者會隨著市場的變動而有變化，跟市場有重大關聯性，因此選擇此參數。
8. **本益比**＝股價/ 每股盈餘EPS
9. 越高：代表目前是以相對較高的價格購入股票，或是要較晚才能收回本金--->市場風險較高。
10. 越低：代表投資者能夠以相對較低價格購入股票，或是能較快回收本金--->市場風險較低。
11. **市值：**指一間上市公司在證券市場上的「市場價格總值」的簡稱。市值的升降反映投資市場的買賣氣氛。
12. **成交值(Sale)：**成交值越大代表看好公司，而成交值與股價的相關係數為正，因程式驗證結果不符合預期，因此選擇不予以保留。
13. **報酬率：**投資淨損 / 總投入資金 \* 100%。一種衡量績效的標準，用於評估投資的效率或營收能力，以及不同投資效率的比較。
14. **週轉率：**一段時間的股票交易數量 / 股票發行數量 \* 100%。指的是一段時間內買賣的比重，也就是一檔股票在一段時間內的發燒、熱絡程度。周轉率高，代表該檔股票買氣高、流通性好、變現能力強 ; 反之，股票的周轉率越低，代表該股票少人關注。
15. **Tobins Q：**公司的市場價值/資產重置成本。大於1時，代表股票市價大於發行成交價，企業家願意投資或增資設廠的意願會提高。反映到股價行為上，當股票價格上漲，發行公司總是在此時機增資發行新股，以籌集資金設廠，因此當 Tobin Q比率提高，代表股票價格上漲，加強了企業投資的興趣。
16. **Tobins Q\_A：**tobins A和上述tobinsq兩者數據幾近相同， 唯一差別在該變數有加特別股本，為了了解是否會影響結果，故我們選擇採用。
17. **負債比率(debtr)：**負債/總資產 \* 100%。當負債比率過於不正常時，代表公司資產不平衡，營運風險增加，因此投資人會考量公司營運狀況，進而評估是否投資，當過多投資人因負債比率過高選擇不投資公司時，股價下跌機率也會隨之增加。
18. **成交值：**成交量 ＊ 股價。可看出市場資金流動方向，投資人因此可跟上趨勢，投資於近期熱門產業。
19. **嘗試模型中變數的敘述統計**
20. **嘗試模型各變數(去除極端值前)**
21. 模型一(全參數，年限由2004開始)



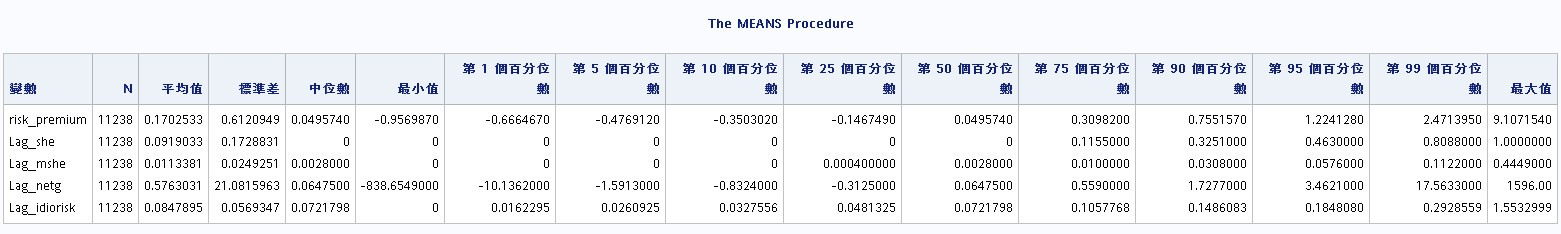
1. 模型二(最佳MAE的參數，年限由2015開始)



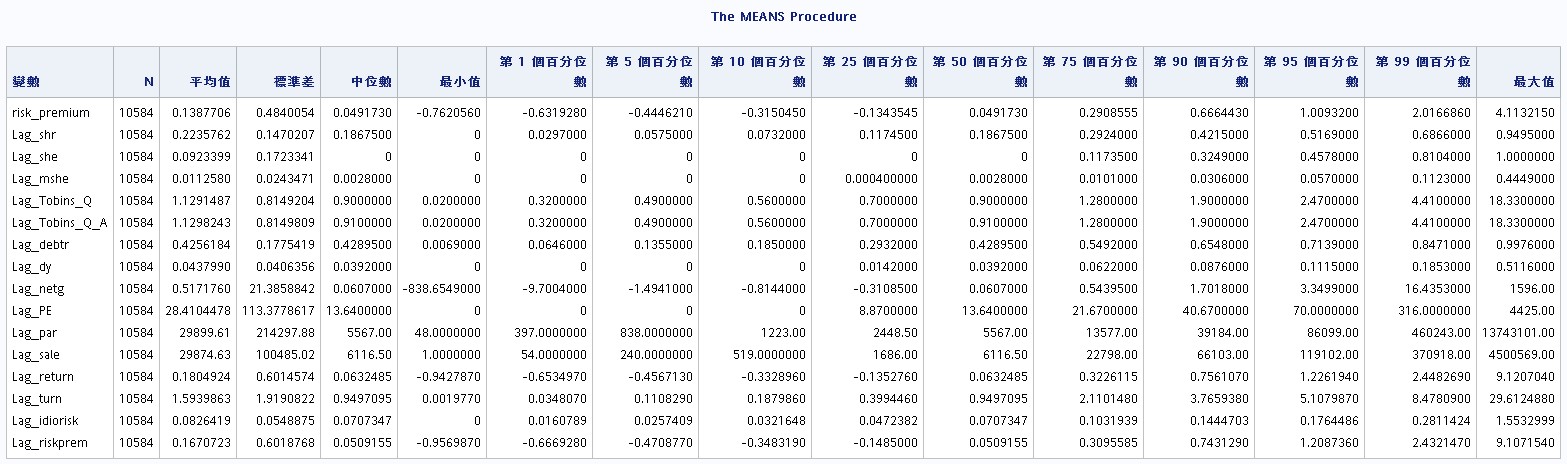
1. 模型三(最佳RMSE的參數，年限由2018開始)



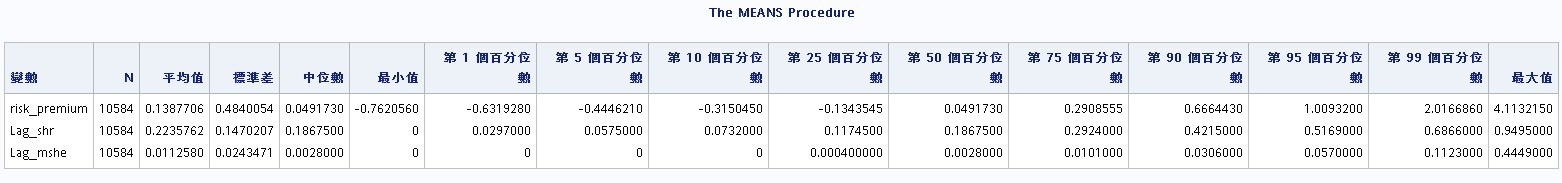
1. 模型四(MAE, RMSE平衡後 較好的參數，年限由2004開始)



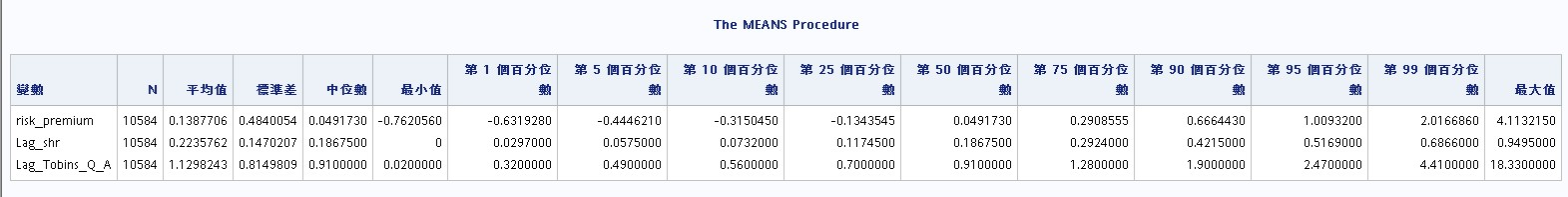
1. **嘗試模型各變數(去除極端值後)**
2. 模型一(全參數，年限由2004開始)



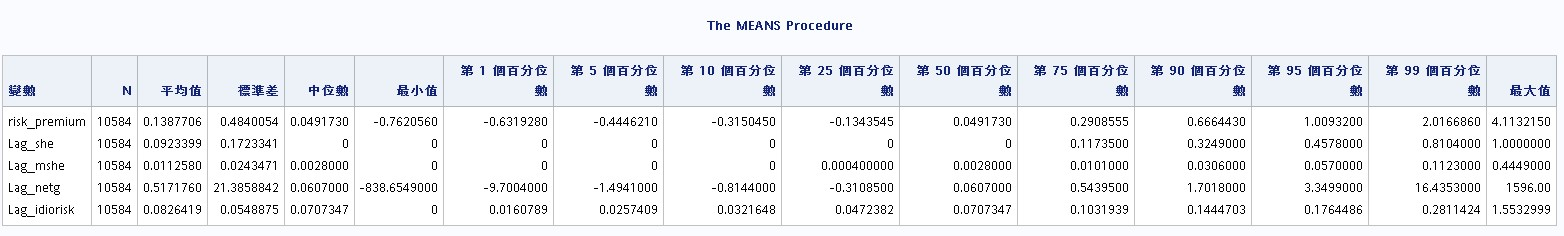
1. 模型二(最佳MAE的參數，年限由2005開始)



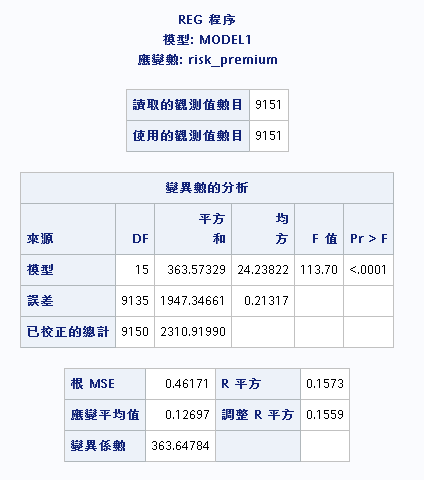
1. 模型三(最佳RMSE的參數，年限由2018開始)

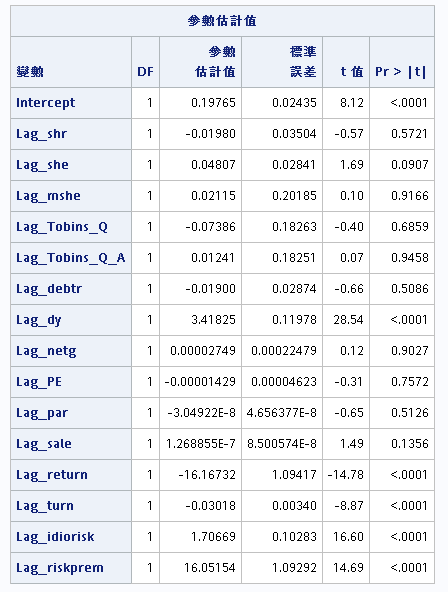


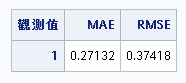
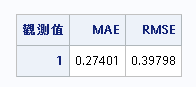
1. 模型四(MAE,RMSE平衡後 較好的參數，年限由2004開始)



1. **嘗試模型之迴歸結果與準確度**
2. **嘗試模型一**： 全參數，年限由2004開始







1. **嘗試模型二**：最佳MAE的參數，年限由2005開始

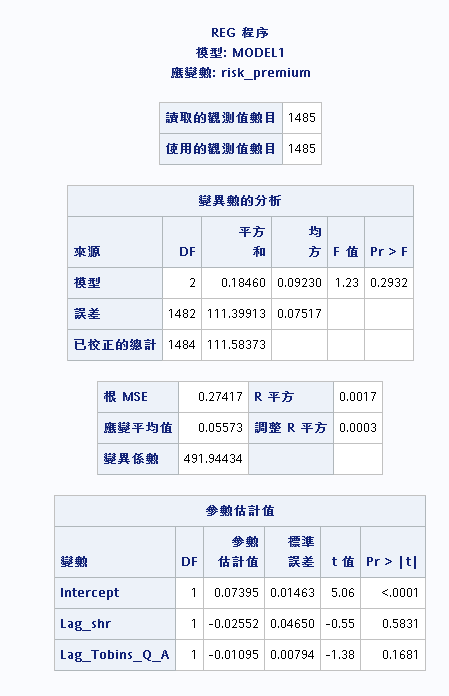
此模型共選取董監質押比率及經經理人質押比率，共兩變數。

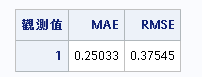


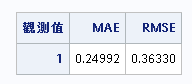




1. **嘗試模型三**：最佳RMSE的參數，年限由2018開始







1. **嘗試模型四**：MAE,RMSE平衡後 較好的參數，年限由2004開始

同最終模型，不重複貼圖