

其他有利審查資料(專題內容為刪減版)

目錄

利用股價預測模型與投資組合管理之多元股票資訊網站	2
摘要	3
研究方法	4
實驗結論	6
中美貿易戰對台灣出口至中國和美國的影響	7
摘要	8
研究方法	9
模型設定	11
結論	12

國立東華大學資訊工程系

National Dong Hwa University

112 學年度大學部畢業專題研究報告

112 CSIE Undergraduate Project Report

利用股價預測模型與投資組合管理之多元股票資訊網站



指導教授 Advisor :

高韓英教授

專題參與人員 Team Member :

鄭羽農

陳仁傑

吳育臻

高鈺雯

中 華 民 國 113 年 5 月 20 日

摘要

本專題針對台灣股票市場，提出一個預測未來一個月個股股價預測模型，此模型結合長短期記憶網絡（LSTM）、自回歸整合移動平均模型（ARIMA）和馬爾可夫鏈蒙特卡羅方法（MCMC）的股價預測模型。通過線性規劃技術，將這些模型組合起來，以最小化均方根誤差（RMSE）為目標，達到最佳預測效果。

另外，針對使用者所輸入之投資組合，也建立一個模型，此模型參考了馬可維茲模型（Markowitz Model），找出最小平方法和最大夏普率之投資組合權重供使用者參考。

本專題最終成果以網站形式呈現，功能包括風險評估、投資小學堂、股票基本面資訊、自選股票管理、系統推薦和投資組合管理，為使用者提供全面的投資建議和預測，並提供一個直觀且易於操作的用戶界面。

關鍵詞：LSTM、時間序列、股價預測、投資組合、股票網站

1、研究方法和結果

1.1 資料處理

本模型使用 Yahoo Finance[17]中所有台灣上市公司之股票自 2020/01/01 以來至今之每日股票交易資料中的收盤價格，並將資料前 80%作為訓練集，後 20%作為測試集。針對資料的處理，在 LSTM 中，將資料標準化處理，將每個特徵的數據縮放到 0 和 1 之間，以防止數據中的特徵值範圍過大影響模型的訓練，進而加快收斂速度，提高模型效能，而對於 ARIMA 和 MCMC 來說，標準化資料並非必要。ARIMA 模型要求數據需為平穩狀態，若數據非平穩應使用差分處理，而對於 MCMC 模型，數據的範圍控制對於機率分布的影響並不大，因此不特別對此二模型資料進行處理。

1.2 模型設定

使用三個模型、LSTM、ARIMA、MCMC，最後建立優化問題來混合三個模型的預測，目標函數是最小化預測與過去四個月歷史數據之間的均方根誤差(MSE)。

1.2.1 LSTM

本模型使用多組輸入一組輸出(many-to-one)的架構，並建立兩個 LSTM 層，為防止過度擬合，皆在該層之後分別應用了 35%和 30%的 dropout 率；輸出層根據 LSTM 層的結果產生一個預測，模型的編譯過程中，本模型使用 ReLU 激勵函數、Adam 優化器和均方誤差(mean squared error)損失函數，同時使用準確性(accuracy)作為評估指標。

1.2.2 ARIMA

在 ARIMA 當中，auto_arima 的參數設定優化的是 lbfgs，全稱是 Limited-memory Broyden-Fletcher-Goldfarb-Shanno 算法。它是一種擬牛頓法(quasi-Newton method)，用於在優化問題中尋找局部最優解。

時間序列的前移驗證(Walk Forward Validation)是一種交叉驗證技術，特別適用於時間序列數據的建模和預測。這種驗證方法將數據集分成訓練集和測試集，並使用滾動的方式進行模型訓練和預測。

以下是時間序列的前移驗證的步驟：

1. **初始化訓練集和測試集**：從時間序列數據的開頭選擇一個固定大小的窗口作為初始訓練集，再選擇一個窗口作為初始測試集。
2. **訓練模型**：使用初始訓練集來訓練模型。
3. **預測下一個時間步**：使用訓練好的模型來預測下一個時間步的值。
4. **更新訓練集和測試集**：將窗口向前移動一步，將先前的測試集的時間步加入到訓練集中，同時將下一個時間步添加到測試集中。

5. **重複步驟 2-4**：重複上述步驟，直到測試集中沒有更多的時間步。
6. **計算模型的性能指標**：使用所有測試集上的預測結果來計算模型的性能指標，例如均方根誤差(RMSE)。

1.2.3 MCMC

在 MCMC 當中，每一個時間步驟中根據幾何布朗運動模型計算股票價格的更新，並使用馬爾可夫鏈蒙特卡洛方法生成隨機數來模擬股票價格的隨機波動。

標準正態分佈的概率密度函數(4.1)：

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(x-u)^2}{2\sigma^2}} \quad (4.1)$$

Black-Scholes 公式中的股票價格變化(4.2)：

$$S_t = S_0 \cdot e^{\left(\left(r - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} \cdot Z_t\right)} \quad (4.2)$$

S_t 是時間 t 的股票價格
 S_0 是初始股票價格
 r 是無風險利率
 σ 是波動率
 Δt 是每個時間步的時間間隔
 Z_t 是標準正態分佈的隨機變數，由 MCMC 方法生成

1.2.4 線性規劃

最後將線性規劃用於在三個不同的時間序列預測模型(LSTM、ARIMA、MCMC)之間分配權重，以最小化均方誤差(MSE)。

1. **定義線性規劃變數**：定義了三個變數，分別代表 LSTM 模型、ARIMA 模型和 MCMC 方法的權重。這些權重變數的值將在求解過程中被優化。
2. **定義目標函數**：將均方誤差的表達式添加到模型中作為目標函數。目標是最小化均方誤差，以使模型的預測與實際數據之間的差異最小化。
3. **添加約束條件**：添加約束條件，確保權重變數的取值合理。其中包括權重之和為 1，以及權重的範圍限制(4.3)。

$$\begin{cases} W_{ARIMA} * S_{AIMA} + W_{LSTM} * S_{LSTM} + W_{MCMC} * S_{MCMC} = \hat{S} \\ W_{ARIMA} + W_{LSTM} + W_{MCMC} = 1 \\ MSE(S, \hat{S}) \rightarrow \min \end{cases} \quad (4.3)$$

其中， S 代表實際預測期間的數據； \hat{S} 代表該期間的預測數據； W 代表每個模型預測的權重。這個模型的輸出是加權的預測數據。

2、實驗結果

我們分別對「1216.TW」、「2002.TW」、「2330.TW」和「2317.TW」進行四個月的回測和一個月的回測來呈現結果（圖 4-1 至圖 4-8），並以表格比較其 RMSE 結果（表 4-1、表 4-2）。

LSTM 和 LP 模型在一個月和四個月內的預測表現良好，ARIMA 模型在一個月和四個月的預測中表現一般，RMSE 值相對較高，MCMC 模型只有在波動性低的股價表現良好，其餘在四個月的預測中情況下表現不佳。

可以發現在四個月的預測底下，透過線性規劃調整權重後的模型預測結果和單用一個模型的 RMSE 平均減少 36%，在一個月的預測，則因為鴻海的近期股價波動劇烈，最終反而使平均 RMSE 上升 38%。

表 4-1 一個月的個別模型預測結果的 RMSE

一個月	1216.TW	2002.TW	2330.TW	2317.TW
ARIMA	1.288	4.763	212.348	38.837
LSTM	0.871	0.515	25.850	5.700
MCMC	0.461	0.458	101.632	22.331
LP	0.415	0.295	19.33	18.47
Reduce RMSE(%)	10%	36%	25%	-224%
Average RMSE	-38			

表 4-2 四個月的個別模型預測結果的 RMSE

四個月	1216.TW	2002.TW	2330.TW	2317.TW
ARIMA	0.801	0.255	30.795	3.4307
LSTM	1.02	0.336	14.795	4.0764
MCMC	5.248	0.811	248.424	59.0331
LP	0.196	0.241	11.324	2.990
Reduce RMSE(%)	81	28	23	13
Average RMSE(%)	36			

國立東華大學經濟學系

National Dong Hwa University

111 學年度大學部畢業專題研究報告

中美貿易戰對台灣出口至中國和美國的影響

The impact of the China-U.S trade war on Taiwan's exports to
China and the U.S



指導教授 Advisor :

林忠樑教授

專題參與人員 Team Member :

鄭羽農

王健綸

賴怡名

中 華 民 國 112 年 1 月 20 日

摘要

本研究探討了中美貿易戰對台灣出口至中國和美國的影響，通過使用 2014 年至 2022 年的出口數據，分析台灣在不同課稅階段的貿易變化。研究採用了**雙重差分法 (DID)** 來比較中美貿易戰前後，台灣對兩國出口額及出口佔比的變化。研究結果顯示，2018 年 9 月和 2019 年 9 月後，台灣部分產業對中國的出口顯著下降，而對美國的出口則上升。這反映了中美貿易戰對台灣在國際貿易中的重要影響，尤其是在特定產業，如機械用具等科技產業。

關鍵詞：中美貿易戰、雙重差分法、出口貿易

1、研究方法

本文主要目的為分別研究中美貿易戰四次不同課稅清單過後，台灣是否成為中美貿易戰之受益國，因此實證模型中的被解釋變數為台灣對各國依HS2k產品分類的貿易額，以運用雙重差分法分析中美貿易戰後台灣對中國與美國的貿易值變化。詳細變數定義如下。

一、出口貿易變數

本文的迴歸模型中，被解釋變數分別為台灣在2014年1月至2022年6月之間出口至中美兩國HS2k的出口貿易值，定義如下：

$valueikt$ ：表示台灣在 t 時間出口 k 產品到 i 國的貿易值，以千美元為單位
 $tradeDependenceikt$ ：在 t 時間，台灣出口 k 產品到 i 國的貿易值佔台灣總出口的百分比

二、國內生產毛額

國內生產毛額是指一個國家或地區在一段時間內(一般情況為一年)生產的最終財以及服務所產生的價值，通常用以衡量一個國家或地區的經濟規模。本文中的國內生產毛額資料取自美國、中國和台灣的政府網站，並以美元為計價單位。

$GDPit$ ：表示 i 國在 t 期之國內生產毛額

三、匯率

匯率定義為兩國貨幣之間兌換的比例，通常會將某一國的貨幣設為基準。本文中的匯率資料取自台灣央行的政府網站，並以美元為計價單位。

$Fxit$ ：表示 i 國貨幣在 t 期和美元之間的匯率

四、美國對中國提高課稅商品的虛擬變數

由於中美貿易戰開始後中美兩國間開始互相提高部分產品的關稅，因此對產品是否課稅進行分類並設立對應的虛擬變數，若中美貿易戰後美國對中國某產品有提高關稅，則虛擬變數 usa_tk 在該產品為1，反之為0。

$usa_tk = 1$, 表示中美貿易戰後美國對中國的 k 產品提高關稅
 $usa_tk = 0$, 表示中美貿易戰後美國對中國的 k 產品沒有提高關稅

五、四次不同時間點的課稅清單之虛擬變數

由於中美貿易戰首次課稅時間為 2018 年 7 月，因此建立一時間虛擬變數 *tradewart*，該變數在 2018 年 7 月之前為 0，2018 年 7 月及之後為 1，之後還有 2018 年 8 月、2018 年 9 月和 2019 年 9 月。

tradewart：表示 *t* 次不同時間點的課稅

六、控制月效果和年效果的虛擬變數

考量出口的淡旺季：

montht：表示在 *t* 月時，虛擬變數為 1，反之為 0

yeart：表示在 *t* 年時，虛擬變數為 1，反之為 0

七、課稅產業的虛擬變數

由於課稅清單為美國的 HS code 八位碼，因此有無被課稅的依據為，四次課稅行動累積的八位碼產業佔所有八位碼產業的百分比【附表二】，其中 $i=1\sim 97$ 個產業， $t=1,2,3,4$ ，分別代表 2018 年 7 月、2018 年 8 月、2018 年 9 月、2019 年 9 月的清單公布時間。若百分比大於 50，則設為有被課稅，記做 1，反之則為 0。

Industryt：表示在 *t* 產業有被美國課稅時，虛擬變數為 1，反之為 0

八、產業分別

由於中國、美國提高關稅的產品不同，因此若中美貿易戰對台灣的出口產生移轉影響，在不同的產業上的衝擊勢必有不同程度的影響，因此將產業進行分類。台灣於 1988 年 1 月 1 日起，採用國際海關理事會所編定的國際商品統一分類制度(HS code)，根據 2017 年中華民國海關進口稅則輸出入貨品分類表以 HS 編碼前兩碼的產品分類，將產業分類為 15 組【附表一】，再於迴歸分析中依各產業分別進行估計。

產業類別以 *Industry_N* 表示，*N* 代表產業 1 至 15。產業類別 01 表示為 *Industry_01*。

2、模型設定

雙重差分法(Difference-in-Difference, DID)

雙重差分法主要用於檢驗一個政策或事件的效果，該方法主要以是否受政策影響來區分實驗組與對照組，並以事件發生時點為界，劃分為事件發生前與發生後。本文中雙重差分法主要用來衡量中美貿易戰前後，台灣出口至中國與美國的出口占比和貿易量是否改變。

$trade_{ikt}$ 代表兩個被解釋變數，分別取自然對數 ln_value_{ikt} 、 $ln_tradeDependence_{ikt}$

加入了 $year_t$ 、 $month_t$ 、 $Industry_t$ 年份、月份、以及產品的固定效果，建立以下模型：

$$ln_trade_{ikt} = \alpha_0 + \alpha_1 USA_tk * tradewart + \alpha_2 GDP_{it} + \alpha_3 GDP_TWt + \alpha_4 Fxit + \alpha_5 year_t + \alpha_6 month_t + \alpha_7 Industry_t + \varepsilon_{ik}$$

模型中 $trade_{ikt}$ 代表三個被解釋變數，分別是 ln_value_{ikt} 、

$ln_tradeDependence_{ikt}$ ，其中 ln_value_{ikt} 為台灣在第 t 時間出口 k 產品到 i 國貿易值的自然對數， $ln_tradeDependence_{ikt}$ 為台灣在第 t 時間出口占比 k 產品到 i 國數量的自然對數的 $tradewart$ 為是否為貿易戰期間的虛擬變數，在本文中為 2018 年 7 月、8 月、9 月和 2019

年 9 月以後為 1，2018 年 7 月、8 月、9 月和 2019 年 9 月以前為 0；

GDP_{it} 為 i 國在 t 年的 GDP， $Fxit$ 為表示 i 國貨幣在 t 期和美元之間的匯率， GDP_TWt 是台灣的 GDP。

3、結論

本文探討提升關稅後美國所產生的貿易移轉觀察台灣是否從中受益，我們使用 2014 年 1 月至 2022 年 6 月的 HS code 2 碼的出口月資料並以四次課稅的時間點為分界點去檢視台灣個別產業的出口變化。

本文的研究結果顯示所有個別產業的迴歸在 0 % 25 % 50 % 的課稅水準虛擬變數設定下，迴歸顯示前兩次的課稅都對台灣的出口以及出口占比皆沒有顯著影響。在所有的產業裡不顯著的產業：動物產品與植物產品以及車輛等運輸設備等等產業，我們推測由於台灣本身對於這些產業並沒有相較於其他國家的優勢所以貿易移轉至其他國家導致台灣的這些產業對於美國的出口以及出口占比沒有顯著的上升改變。而對於台灣主要出口的產業機械用具等科技類型的產業就有產生很顯著的美國貿易移轉現象導致台灣對美國的出口以及出口占比有顯著的提升。

2018/7 和 2018/8 月的課稅對台灣出口至美國和中國尚未有顯著性，可能原因是因為課稅的規模和金額並沒有特別大，以及台灣的出口至美國和中國還尚未反應。

中美貿易戰對台灣出口至美國有較多產業呈現顯著的正面效果，尤其食品、機械用具塑膠製品等上升的較為明顯，而對中國則是呈現負面的效果，其中紡織、貴金屬、石料等製品下降的較為明顯。

最後本文的研究結果可以得知在 2018/9 月和 2019/9 的課稅，使

台灣對中國和美國的出口變化較有影響，台灣特定產業對中國的出口相對下降，對美國的出口則是相對上升。