

FCU



ePaper

逢甲大學學生報告 ePaper

**預測台灣蛋價的模型實作：**

**時間序列分析與類神經網路**

**Model Implementation for Forecasting Egg Prices  
in Taiwan: Time Series Analysis and Neural Networks**

作者：林宏展、林長毅、林紘郁、余俊傑

系級：經濟四甲、工業四甲、經濟二乙、資訊二丙

學號：D0837808、D0842954、D1076332、D1031329

開課老師：何思賢教授

課程名稱：資料科學導論：Python 實踐

開課系所：經濟學系

開課學年：111 學年度 第 2 學期

## 中文摘要

雞蛋在臺灣人的日常飲食中扮演著重要的營養供應角色，因此，維持雞蛋價格的穩定性相當重要。本研究從多個來源蒐集 2010 年 11 月 1 日至 2023 年 4 月 30 日之蛋價資料，缺漏部分以插值法補足。我們蒐集影響蛋價的相關因子，建立三個蛋價預測模型：(1) 多元迴歸 (Multiple Regression)，(2) ARIMAX (Autoregressive Integrated Moving Average with Explanatory Variables Algorithm)，(3) 長短期記憶 (Long Short Short-Term Memory, LSTM) 之類神經網路模型。我們以標準的機器學習方法進行訓練和預測，結果顯示，在多元迴歸、ARIMAX 和 LSTM 模型預測之均方根差 (Root-Mean-Square Error, RMSE)，分別為 9.2412、9.3654 和 3.0985 元。因此，LSTM 表現遠遠優於另外兩者，較能捕捉 2023 年台灣較為異常且漲幅較大的蛋價。本研究提供一個且可行的 LSTM 的架構，供後續研究人員或是政府單位建立模型之參考，並提早制定相關措施。

**關鍵字：**

**台灣蛋價、多元迴歸模型、ARIMAX 模型、長短期記憶模型、價格預測**

## Abstract

Eggs assume a significant role in the daily dietary intake of the Taiwanese population, emphasizing their crucial nutritional contribution. Consequently, maintaining stable egg prices holds substantial importance. This study amalgamated egg price data from diverse sources spanning November 1, 2010, to April 30, 2023, with interpolated values to address gaps. Pertinent factors affecting egg prices were collected to establish three egg price forecasting models: (1) Multiple Regression, (2) ARIMAX (Autoregressive Integrated Moving Average with Explanatory Variables Algorithm), and (3) Long Short-Term Memory (LSTM) neural network model. Training and prediction were conducted employing standard machine learning methodologies. The outcomes revealed root-mean-square errors (RMSE) for egg price prediction in multiple regression, ARIMAX, and LSTM models as 9.2412, 9.3654, and 3.0985 units, respectively. Consequently, LSTM outperformed the other two models, effectively capturing the pronounced and larger-scale egg price anomalies expected in Taiwan for the year 2023. This research furnishes a practicable LSTM framework, serving as a reference for future researchers or governmental entities aiming to construct models and implement anticipatory measures.

### **Keyword :**

**Taiwan Egg Prices, Multiple Regression Model, ARIMAX Model,  
Long Short-Term Memory Model, Price Forecasting**

## 目 次

中文摘要.....	1
Abstract.....	2
目次.....	3
一、 研究動機與目的.....	4
二、 文獻探討.....	8
三、 變數定義與資料預處理.....	10
四、 Multiple Regression Model.....	18
五、 ARIMAX Model.....	22
六、 Long Short-Term Memory (LSTM) Model .....	27
七、 結論與未來展望.....	34
參考資料.....	37



## 一、 研究動機與目的

雞蛋對於臺灣人來說具有重要性，不僅是營養豐富的食物，提供了許多美味且經濟實惠的飲食選擇，滿足了臺灣人對於多樣化飲食的需求，且在臺灣市場上相對廉價，便於取得，因此，它成為許多家庭烹飪的主要食材之一。

因此，維持雞蛋價格的穩定性將變成非常重要，穩定的雞蛋價格有助於保護消費者的權益，確保消費者能夠以合理的價格購買到必需的食物。另外，穩定的雞蛋價格對於農戶來說也是至關重要的，農戶在生產雞蛋時需要投入資源，如飼料、飼養成本等，維持雞蛋價格的穩定性能夠提供農戶更穩定的收入，鼓勵他們進一步投資和提高生產效率，並有助於促進供應鏈的協調和市場的效率，提供供應商和零售商可靠的市場信號，促進市場的順暢運作。此外，雞蛋作為臺灣農業的重要部分，其價格波動可能對整體經濟產生波及效應，如通脹壓力、消費者支出變動等，維持雞蛋價格的穩定性有助於促進農業部門的可持續發展，並維護整體經濟的穩定。

然而，如圖 1.1 所示，過往臺灣有經歷過幾次蛋價漲幅較高的時期，如 2015 年開始至 2016 年 6 月，相關新聞報導截圖如圖 1.2 所示，接著，在 2018 年 9 月至 2019 年 5 月也經歷了另一波的漲價，相關新聞報導截圖如圖 1.3 所示。甚至在近期，從 2022 年開始蛋價一直不斷地攀升，並在 2023 年到了歷史新高，有史以來最高的蛋價，加上蛋價仍不斷的上漲，造成民眾搶蛋，各大超市蛋架上的雞蛋常常被一掃而空，許多人買不到蛋，又可能進一步漲價。本研究將既有模式問題整理為以下兩點：

### 1. 蛋價易受各種因素影響，造成民眾搶購：

蛋雞容易受氣溫、禽流感 and 天災等因素而造成產蛋量下降，甚至是雞隻死亡，此外供應蛋雞的飼料主要為玉米和小麥，大多需從國外進口，飼料進口價格常常浮動不定，加上近期烏俄戰爭導致玉米小麥價格上漲，也直接或間接的反應在蛋價上。最後蛋價上漲導致民眾恐慌而搶購，讓蛋價進一步上漲。

## 2. 目前較無針對未來蛋價的可靠預測模型

目前僅有各大農委會或是政府的相關機關網站，有整理每日或是每週的蛋價、飼料價格和蛋雞隻數等會影響蛋價的資料，但目前並無整合的網站，可以同時一次看到多維度的資料，造成民眾或是飼養人員查找的不便，且目前都僅以相關的資料表格或是敘述性統計的折線圖呈現資料，查看者僅能看到過往的趨勢和目前的資料，並無相關的蛋價預測模型，能夠預測未來有可能的上漲時間，並且提早準備和預防。

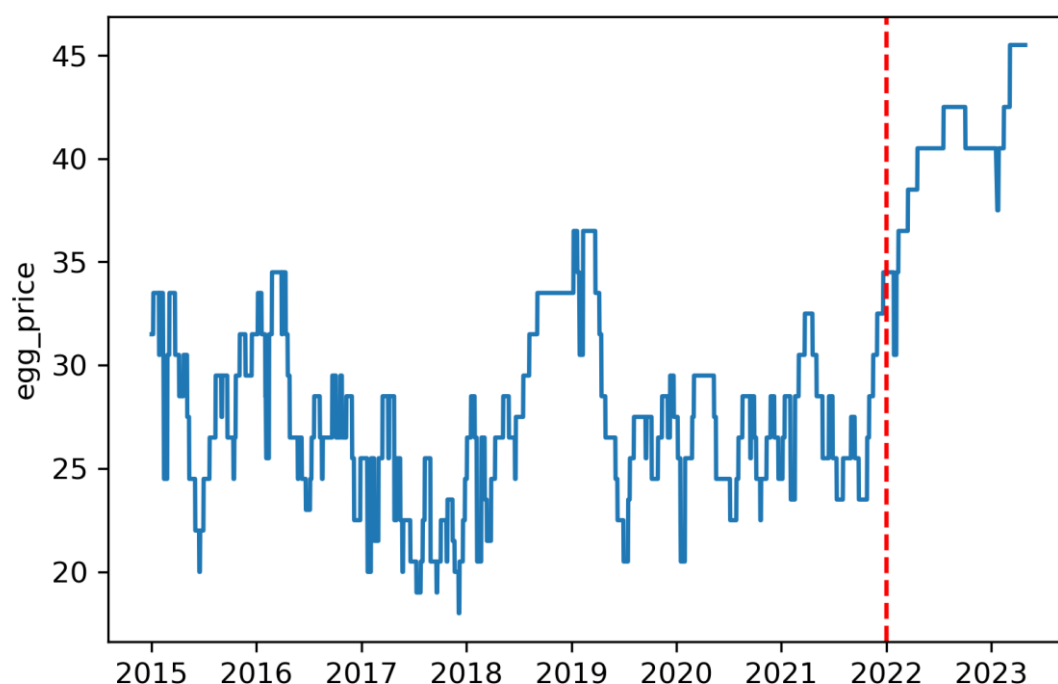


圖 1.1、臺灣歷史蛋價折線圖



圖 1.2、2015 年開始至 2016 年 6 月蛋價相關報導



圖 1.2、2018 年 9 月至 2019 年 5 月蛋價相關報導

綜上所述，本研究期望透過參考過往蛋價漲價的相關報導，以蒐集可能影響蛋價的相關因子，並參考預測蛋價的相關文獻，建立預測蛋價時間序列的模型，接著輸入可能影響蛋價的因子資料，最後評估模型建立之正確性和精確性，並將預測結果供一般民眾和政府機關參考。本研究期望模式歸納於以下兩點：

**1. 蒐集可能影響蛋價的相關因子，並建立蛋價資料集：**

透過查找相關的新聞報導，以收集可能影響蛋價的因子，以此做為資料及的欄位拓展資料集的維度，接著依照可能影響蛋價的因子進行資料探勘，到民間農業團體、政府機關和公開資料集等相關網站，查找是否有影響蛋價因子的歷史資料，最後將進行補值和去除異常值等資料預處理，完成蛋價資料集，並提供後續模型建立使用。

**2. 建立多個蛋價預測模型，並比較個模型之間的優劣：**

透過參考過往預測價格模型，並評估各模型對於預測蛋價的可行性與準確定，找到對於預測蛋價較為可行的演算法，接著透過建立多個模型，比較各模型是否能藉由歷史資料預估未來蛋價，取代人工預測的方式，最後以均方根誤差評估各模型的優劣，並以最佳預測模型作為本研究之解決方案。





## 二、 文獻探討

### ➤ Multiple Regression Model

房地產價格於建築、投資、公益等相關議題會有重大影響，如何建構精確預測房地產價格的模型，一直是一個具有挑戰性的課題，因此 [Qingqi Zhang\(2021\)](#) 對基於多元線性回歸房價預測的廣泛研究，將重點放在對模型影響最大的關鍵因素上，評估具有特定特徵的房地產的經濟價值，並確定回歸因素背後的因果關係。首先該研究利用 Spearman 相關係數分析影響房價的主要因素，選取影響一般房價的顯著因素，進行組合分析演算法。接著，該研究建立房價預測的多元線性回歸模型，並應用波士頓房地產價格數據集對該方法進行了檢驗，通過本文的數據分析和測試，綜上所述，多元線性回歸模型可以在一定程度上有效地預測和分析房價，而演算法仍可以通過更先進的機器學習方法進行改進。

### ➤ ARIMAX Model

雞蛋在國人飲食佔有相當重要的成分，而近年來雞蛋價格屢創新高且波動大，導致許多業者與消費者對於高價雞蛋有苦難言。因此，[唐婉庭\(2019\)](#) 想藉由觀察過去雞蛋價格，與蛋雞內部產業之生產週期變化，以及外部環境因素，加以探討影響雞蛋產地價格變動之因素，數據蒐集自 2005 年 1 月至 2017 年 12 月之數據，作為執行 2018 年雞蛋產地價格之預測依據，並採用時間序列 ARMA 模型，進行設計基本模型以及加入內外部因素後的延伸模型，利用兩種不同的模型架構對月平均雞蛋產地價格走勢之檢定，最後比較不同預測模型之準確度後建立雞蛋產地預測模型，以確認影響雞蛋產地價格變動之主要因素。結果顯示，要預測臺灣雞蛋月平均產地價格，不僅需考慮過去雞蛋產地價格，也須觀察其他變數。影響臺灣雞蛋產地價格變動的因素有：前兩期的產地價格、前兩期的產蛋箱數、當期換羽隻數、當期淘汰隻數、前一期產地氣溫以及農曆新年因素所影響。因此，相關單位必須整體追蹤監控上述數據資訊之變動，如此才能有效降低雞蛋產地價格無預警波動以穩定產銷平衡。

### ➤ Long Short-Term Memory (LSTM) Model

時間序列的問題，是在機器學習的知識領域中，應用和需求最廣泛的問題之一，因此，王柏鈞 (2020) 透過 LSTM 進行時間序列分析以股價預測，首先該文章透過 Sliding window 技巧將較長的時序資料，變成許多較短的時序資料，利用將 Stride (步伐) 設定為 1，而 window-size (窗口大小) 設定為 7，讓資料增加為 7 倍，相當於每隔一天就取一個星期的資料，接著將透過 Sliding window 處理後的股價資料，輸入至 LSTM 模型中進行訓練，但該文章發現今日股價與昨日收盤價較為相關，導致 LSTM 模型容易 OverFitting，訓練效果不佳，因此透過在層與層之間加入 20% 的 Dropout，最後成功建構預測股價之 LSTM 模型，均方根誤差 (Root-Mean-Square Error, RMSE) 為 0.69。



### 三、 變數定義與資料預處理

本研究以文獻之變數選擇作為參考，並加入飼料價格、氣溫平方項、食物類消費者物價指數三變數，本研究預測氣溫應有最適溫度，因此加入氣溫平方項。資料期間為 2010 年 11 月 1 日至 2023 年 4 月 30 日，本研究將訓練集定為 2010 年 11 月 1 日至 2020 年 10 月 28 日之資料（佔 80%），測試集為 2022 年 10 月 29 日至 2023 年 4 月 30 日（佔 20%），各變數定義如表 3.1 所示：



表 3.1、變數名稱與來源

變數名稱	定義與來源
雞蛋價格	以行政院農委會資料開放平台提供之家禽交易行情（白肉雞/雞蛋）作為蛋價（日資料）。
產蛋箱數	當週平均每日雞蛋生產量（箱/200 顆），以中華民國養雞協會提供之歷年蛋雞生產數量統計總表，表中之產蛋隻數（月資料）。
淘汰隻數	各蛋雞場淘汰蛋雞數量之總和，以中華民國養雞協會提供之歷年蛋雞生產數量統計總表，表中之淘汰隻數（月資料）。
換羽隻數	各蛋雞場執行換羽之蛋雞數量總和，以中華民國養雞協會提供之歷年蛋雞生產數量統計總表，表中之目前換羽隻數（月資料）。
飼料價格	以中央畜產會畜產品查詢系統提供之現行品項平均價格查詢的產地農場價格，其中的蛋雞飼料作為飼料價格（月資料）。
氣溫	以農業氣象觀測網監測系統提供之農業站每日氣溫數據，採用 6 個蛋雞主要生產地區做平均，分別為：彰化、嘉義、高雄、屏東、台南、雲林，對應之農業站分別為：台中農改、義竹分場、旗南農改、高雄農改、畜試所、雲林分場（日資料）。
氣溫平方項	為捕捉最適氣溫。
農曆新年	虛擬變數，春節第一天所在之月份，該月份值為 1，其餘則為 0。
食物類消費者物價指數（CPI）	以中華民國統計資訊網提供之食物類消費者物價指數（月資料）。

資料大都為月資料，因此本研究使用內插法進行補值，各變數補值方法如表 3.2 所示：

表 3.2、各變數補值與處理

變數名稱	補值與處理方法
雞蛋價格	當天蛋價為「議價」、「休市」，則以前後一天價格平均代替，若需連續補多天價格，則須補值的第一天補前一天價格，最後一天補後一天之價格，中間以第一天與最後一天之平均代替。
蛋雞數量	先將月資料除以 30 天，再以三次樣條插值法（Cubic spline interpolation）補完整個月的值。
產蛋箱數	先將月資料除以 30 天，再以三次樣條插值法補完整個月的值。
淘汰隻數	先將月資料除以 30 天，再以三次樣條插值法補完整個月的值。
換羽隻數	先將月資料除以 30 天，再以三次樣條插值法補完整個月的值。
飼料價格	使用線性插值法（Linear interpolation）補完整個月的值。
氣溫	使用線性插值法對缺漏值進行補值。
消費者物價指數	將月資料補成每日資料。

插值法程式碼示例如圖 3.1 和圖 3.2 所示：

```
start_date = pd.to_datetime('2005/01/01')
end_date = pd.to_datetime('2023/04/30')
daily_dates = pd.date_range(start=start_date, end=end_date, freq='D')
# 蛋箱數
cs = CubicSpline(eliminate['date'].astype(np.int64), eliminate['box_egg'])
box_egg = cs(daily_dates.astype(np.int64))
```

圖 3.1、三次樣條插值法

```
feed = pd.read_excel('feed2.xlsx')
feed['y'] = feed['date'].apply(lambda x :int(x.split('年')[0])+1911)
feed['m'] = feed['date'].apply(lambda x :x.split('年')[1].replace('月',''))
feed['date'] = pd.to_datetime(feed['y'].astype(str) + '/' + feed['m'].astype(st
feed2 = pd.DataFrame()
start_date = pd.to_datetime('2005/01/01')
end_date = pd.to_datetime('2023/04/30')
daily_dates = pd.date_range(start=start_date, end=end_date, freq='D')
feed2['date'] = daily_dates.strftime('%Y/%m/%d')
feed2 = feed2.merge(feed, on='date', how='left')
feed2 = feed2.drop(feed2.iloc[:,[3,4]], axis=1)
feed2['feed'] = feed2['feed'].interpolate(method='linear')
```

圖 3.2、線性插值法

各變數補值後之結果如圖 3.3、圖 3.4、圖 3.5、圖 3.6、圖 3.7 和圖 3.8 所示：

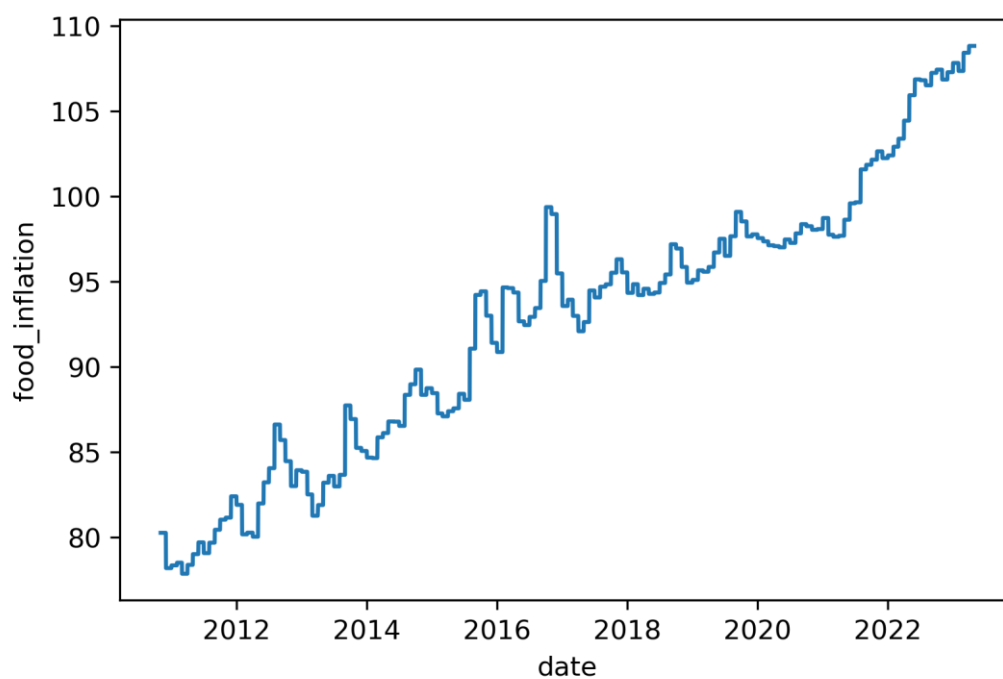


圖 3.3、食物類 CPI 補值結果

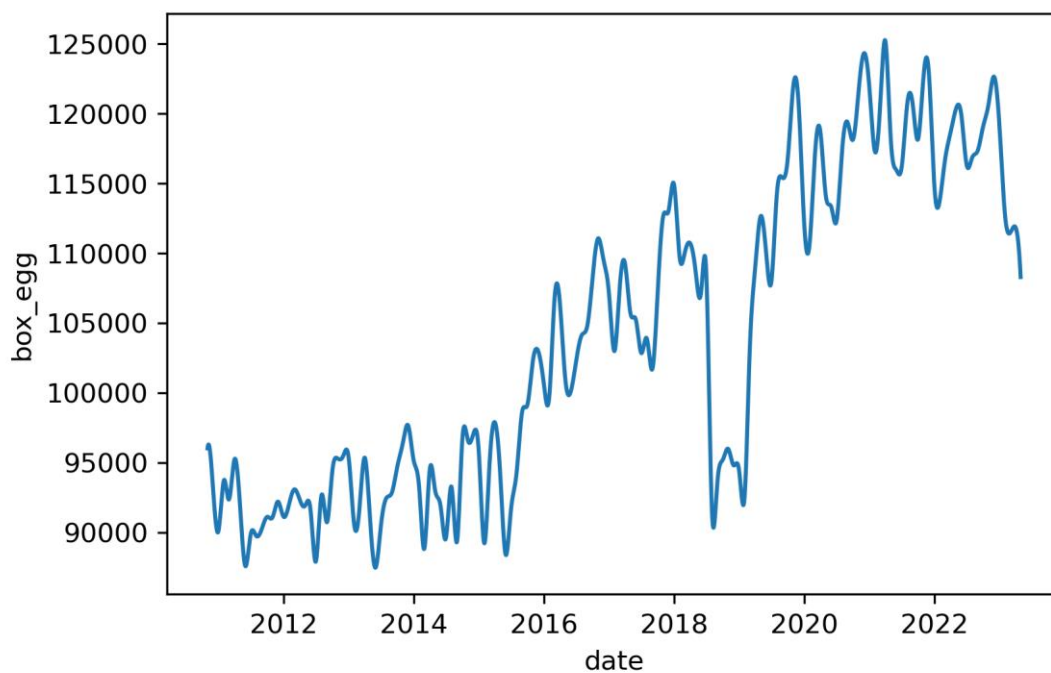


圖 3.4、產蛋箱數補值結果

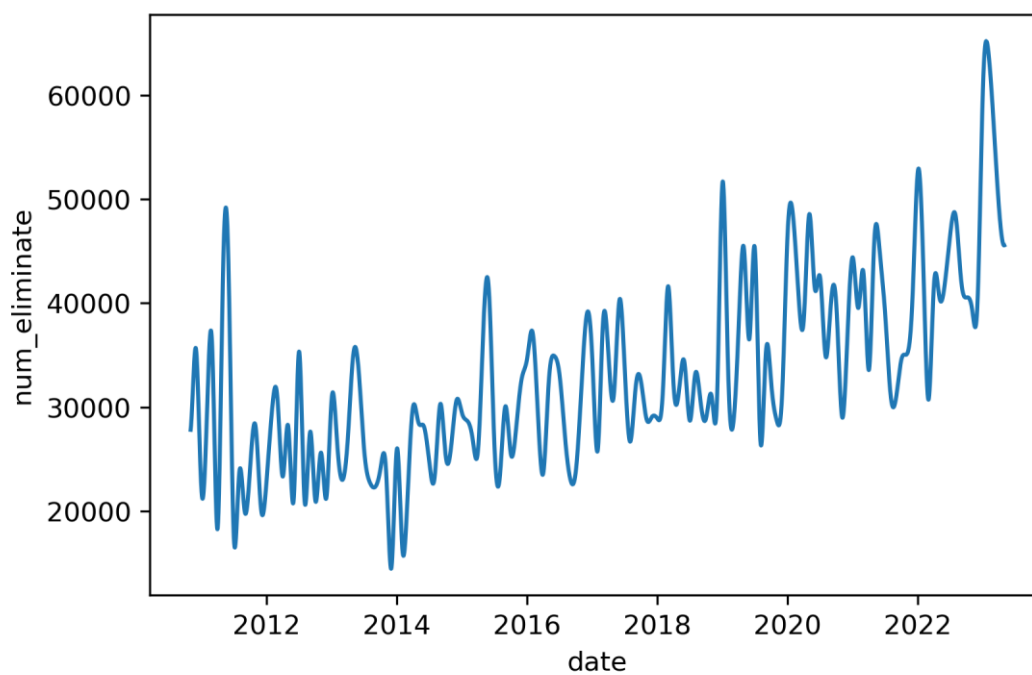


圖 3.5、淘汰隻數補值結果

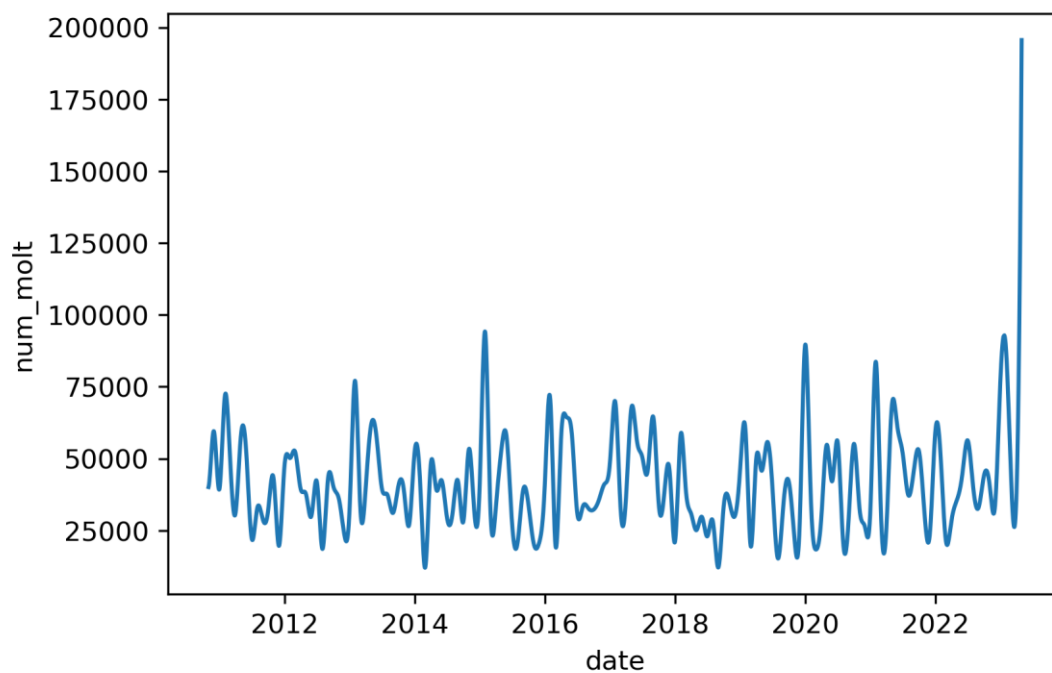


圖 3.6、換羽隻數補值結果

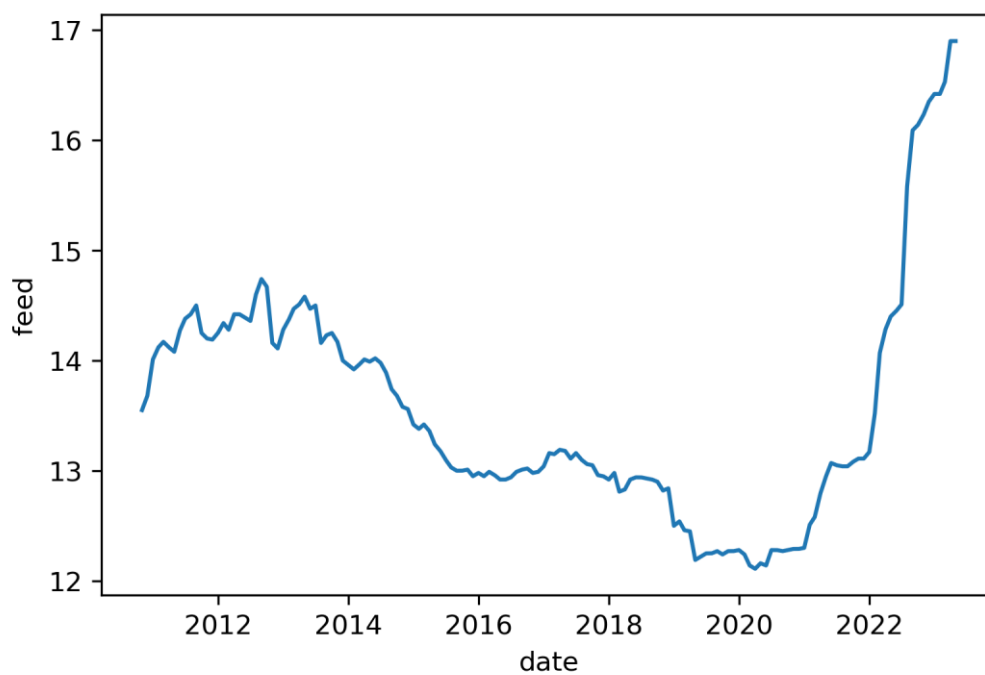


圖 3.7、飼料價格補值結果



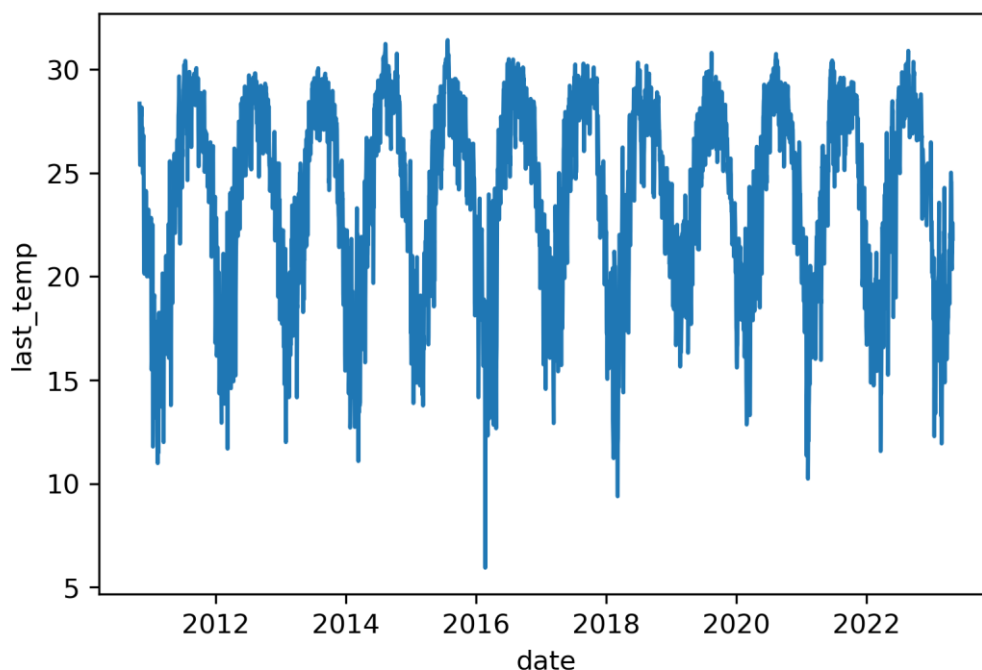


圖 3.8、氣溫補值結果

本研究透過表 3.3 初步觀察資料之敘述性統計和圖 3.9 觀察各變數之相關性。

表 3.3、敘述統計表

變數	觀察值	平均數	標準差	最小值	最大值
農曆新年(是=1)	4559	.082	.275	0	1
蛋價	4559	28.456	5.629	16	45.5
食物類 CPI	4559	92.544	7.992	77.86	108.81
產蛋箱數	4559	103908	11113.6	87454.7	125250
淘汰隻數	4559	32943.2	8615.46	14457.1	65204.7
換羽隻數	4559	41487.5	16156	12000.5	195604
飼料價格	4559	13.533	1.029	12.11	16.9
前一天氣溫	4559	23.988	4.51	5.933	31.417

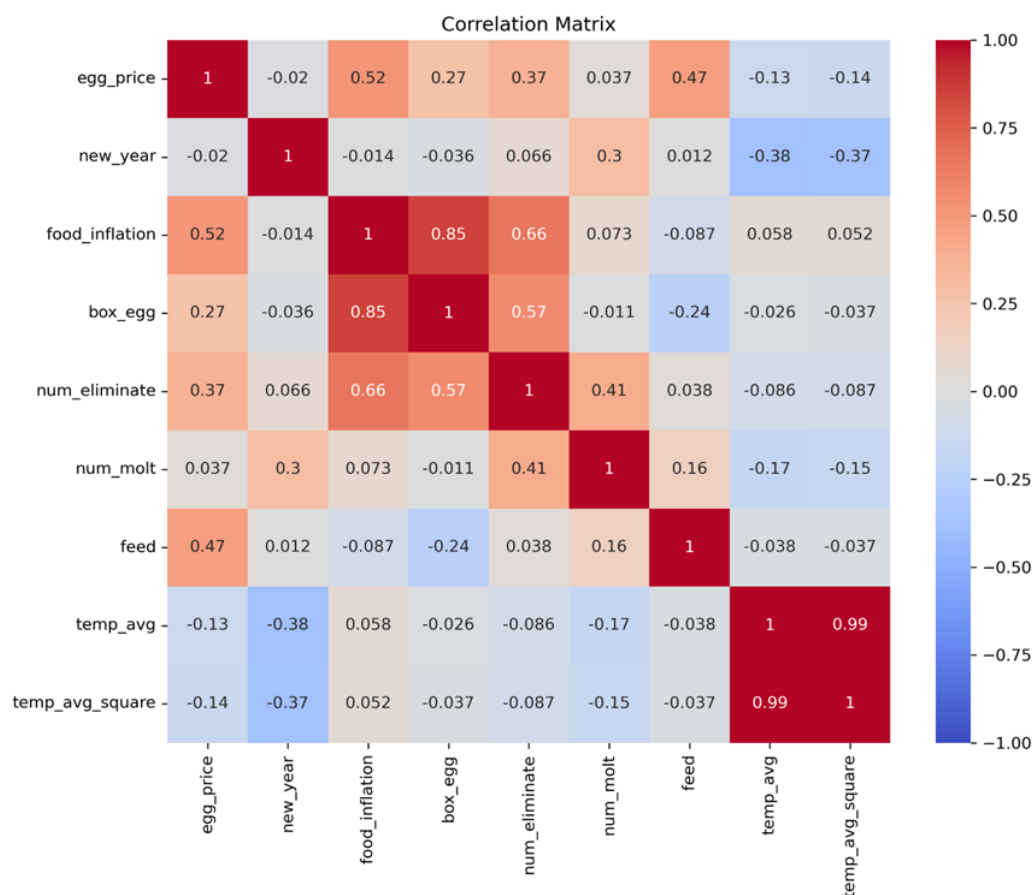


圖 3.9、各變數相關係數表

為使模型結果不受單位影響，本研究將所有變數正規化（Normalization）後再使用，正規化結果如圖 3.10 所示。

	0	1	2	3	4	5	6
0	0.0433912	0.0759289	0.159256	0.0875861	0.313425	0.275372	0.419461
1	0.0433912	0.0759289	0.162553	0.0927788	0.311215	0.276382	0.419851
2	0.0433912	0.0759289	0.166035	0.0982517	0.308665	0.277392	0.42024
3	0.0433912	0.0759289	0.169677	0.103972	0.305822	0.278403	0.42063
4	0.0433912	0.0759289	0.173454	0.109907	0.302732	0.279413	0.421019
5	0.0433912	0.0759289	0.177344	0.116024	0.299442	0.280423	0.421409
6	0.0433912	0.0759289	0.181321	0.12229	0.295998	0.281433	0.421798
7	0.0547397	0.0775444	0.241653	0.226226	0.263322	0.300626	0.429199
8	0.0547397	0.0775444	0.242351	0.228619	0.266566	0.301531	0.42876
9	0.0547397	0.0775444	0.242675	0.230525	0.270525	0.302436	0.428321

圖 3.10、資料正規化之部分結果

## 四、 Multiple Regression Model

如圖 4.1 所示，為本研究 Multiple Regression Model 之程式碼。

```
# 提取目標變數
y = df['egg_price']
# 提取外生變數
X = df.drop(['date', 'egg_price', 'full_inflation', 'num_chicken', 'corn', 'current_temp'], axis=1)
X = np.array(X)
# 加入截距項
X = sm.add_constant(X)

# 測試集大小
test_size = int(len(X) * 0.2)

train_x, test_x = X[:-test_size], X[-test_size:]
train_y, test_y = y[:-test_size], y[-test_size:]

model = sm.OLS(train_y, train_x)
model_fit = model.fit()

pred = model_fit.predict(test_x)
# 計算rmse、r平方
mse = np.mean((pred-test_y)**2)
rmse = np.sqrt(mse)
model_fit.summary()
r2 = model_fit.rsquared
```

圖 4.1、多元迴歸模型程式碼

多元迴歸模型以線性關係解釋應變數與自變數之間如何互相影響，如表 4.1 所示，若逢農曆新年（虛擬變數 = 1），則蛋價會下降 2.0166 元；食物類 CPI 上升 1 單位，蛋價則上升 0.5339 元，以此類推。

若逢農曆新年、產蛋箱數越多、淘汰隻數越多、換羽隻數越多，蛋價會越低；食物類 CPI 越高、前一期氣溫越高，蛋價則越高，與參考文獻之研究結果相同，RMSE 為 9.2412，R 平方為 0.313。

本研究預期氣溫平方項應為負向，如此才會有最適氣溫，但結果與預期相反，本研究猜測與淘汰隻數有關，氣溫越極端，淘汰隻數越多，氣溫的一部份已被淘汰隻數解釋，而導致結果與預期相反。

表 4.1、多元迴歸模型結果

	coef	std err	t	P> t
const	18.0286	4.965	3.631	0.000
new_year	-2.0166	0.234	-8.610	0.000
food_inflation	0.5339	0.023	23.159	0.000
box_egg	-0.0003	1.1e-05	-29.006	0.000
num_eliminate	-9.407e-05	1.15e-05	-8.180	0.000
num_molt	-5.161e-05	4.69e-06	-11.008	0.000
feed	-0.1152	0.209	-0.551	0.582
last_temp	0.4161	0.126	3.307	0.001
last_temp_squared	-0.0173	0.003	-6.306	0.000

如圖 4.2 所示，多元迴歸模型之預測能力不甚理想，RMSE 為 9.2412，R 平方為 0.313，與實際價格平均相差大約 9 元。

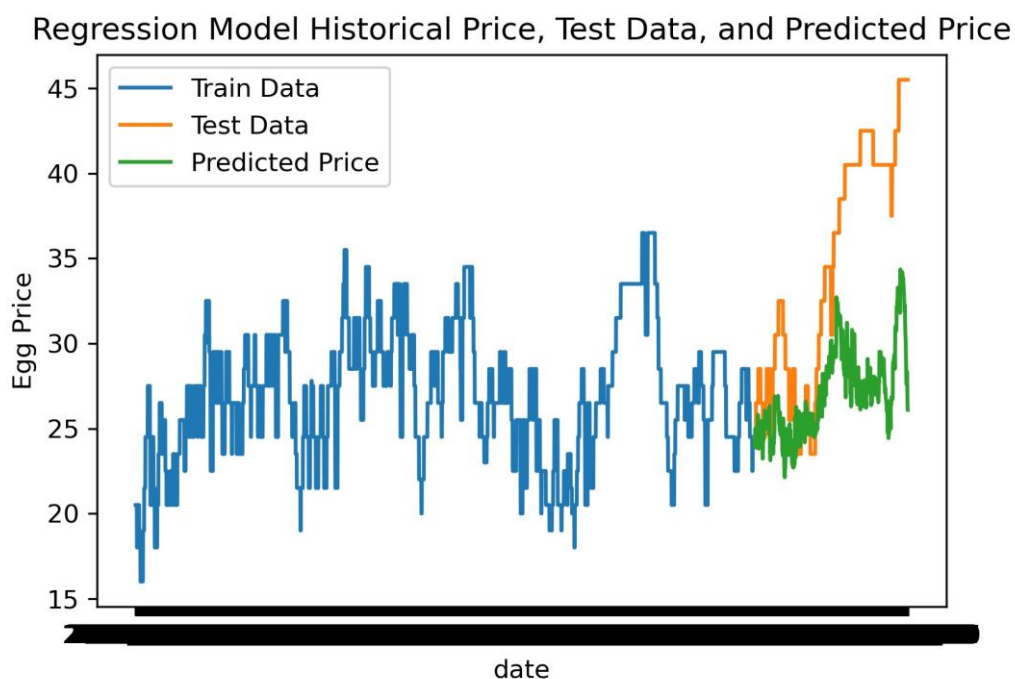


圖 4.2、多元迴歸預測結果

由表 4.2 可知，將氣溫變數去除後，除了新年變數變得不顯著之外，各變數結果變化不大，且 RMSE 上升至 11.7518，R 平方下降至 0.179，代表氣溫對於預測蛋價依舊是重要變數。

表 4.2、多元迴歸模型結果（去除氣溫）

	coef	std err	t	P> t
const	37.6697	5.194	7.252	0.000
new_year	-0.0608	0.237	-0.256	0.798
food_inflation	0.3512	0.024	14.614	0.000
box_egg	-0.0003	1.19e-05	-23.184	0.000
num_eliminate	-6.464e-05	1.24e-05	-5.195	0.000
num_molt	-3.342e-05	5.07e-06	-6.596	0.000
feed	-0.8400	0.226	-3.709	0.000

表 4.3 為將訓練集拉長至 2022/01/26 (佔 90%)，飼料價格方向改變並符合我們的預期，其餘變數方向均無變化，RMSE 下降至 6.4765，準確度大幅上升，表示多元迴歸模型要以 2022 年之前的資料，預測 2022 後的特殊情形較困難。

表 4.3、多元迴歸結果（訓練集拉長）

	coef	std err	t	P> t
const	-14.6166	3.497	-4.180	0.000
new_year	-2.1035	0.223	-9.434	0.000
food_inflation	0.6141	0.019	32.182	0.000
box_egg	-0.0003	9.36e-06	-28.429	0.000
num_eliminate	-3.667e-05	1.03e-05	-3.575	0.000
num_molt	-5.673e-05	4.29e-06	-13.238	0.000
feed	1.2265	0.148	8.271	0.000
last_temp	0.5246	0.119	4.426	0.000
last_temp_squared	-0.0204	0.003	-7.878	0.000

有鑑於食物通膨 (food\_inflation) 和產蛋箱數 (box\_egg) 兩因子可能需要一段時間才會反映在蛋價上，而非直接對當日蛋價產生立即性的影響，因此本研究將食物通膨因子延後 30 天，產蛋箱數因子延後 7 天，也就是透過 30 天後的食物通膨因子和 7 天後產蛋箱數因子預測當日的蛋價，其餘因子根據唐婉庭 (2019) 使用當天資料預測，以測試是否能提升最終的預測結果，多元迴歸模型結果如表 4.4 所示，多元迴歸預測結果如圖 4.3 所示，RMSE 為 13.2135，R 平方為 0.257，與實際價格平均相差大約 13 元，預測結果誤差不減反增。

表 4.4、多元迴歸模型結果（食物通膨和產蛋箱數因子延後）

	coef	std err	t	P> t
const	39.8015	5.030	7.913	0.000
new_year	-1.7780	0.239	-7.426	0.000
food_inflation	0.3233	0.024	13.417	0.000
box_egg	-0.0003	1.15e-05	-26.848	0.000
num_eliminate	-6.282e-05	1.18e-05	-5.315	0.000
num_molt	-3.187e-05	4.83e-06	-6.602	0.000
feed	-1.3288	0.218	-6.088	0.000
temp_avg	1.2518	0.130	9.649	0.000
temp_avg_square	-0.0332	0.003	-11.729	0.000

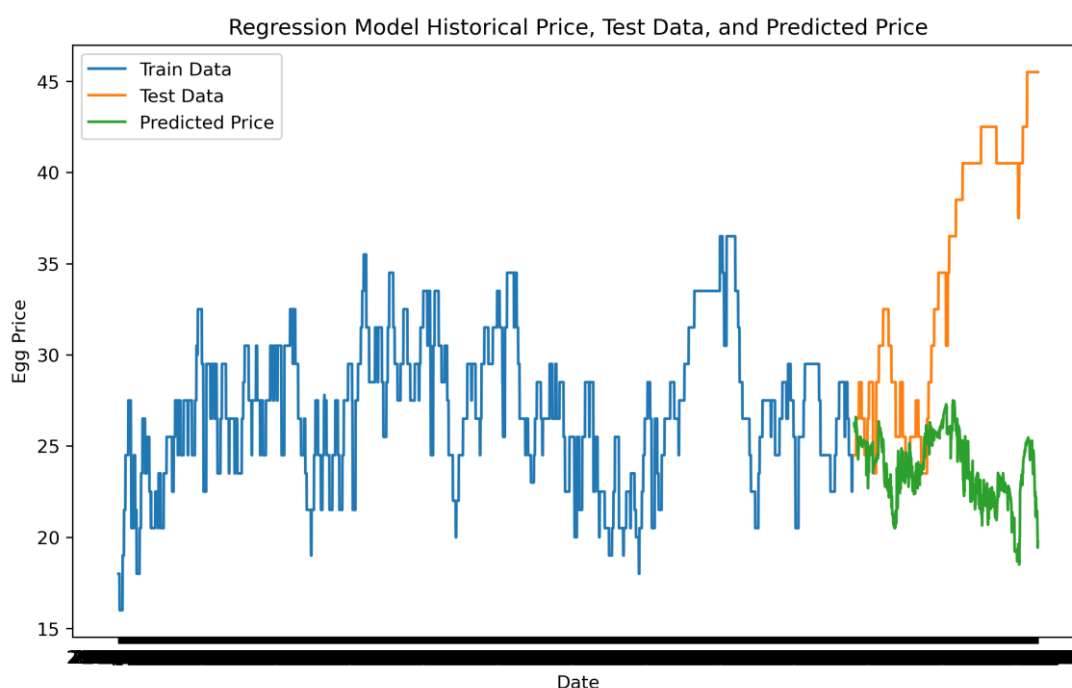


圖 4.2、多元迴歸模型預測結果（食物通膨和產蛋箱數因子延後）

## 五、 ARIMAX Model

自我迴歸-移動平均差分模型 (ARIMAX) 為一時間序列模型，自我迴歸 (AR) 項為使用第  $t$  期之應變數作為解釋變數放入模型，負責捕捉重點期間，差分 (I) 為將前或後  $t$  期減去當期，以穩定時間序列，平均移動 (MA) 項為以前  $t$  期作為解釋變數放入模型，負責捕捉長期趨勢，ARIMAX 為 ARIMA 之延伸模型，X 代表允許納入外生變數。

本研究通過穩定性檢定、自我相關函數圖 (ACF)、偏自我相關函數圖 (PACF)、季節性檢定來選擇模型參數，以下為各檢定之結果：

穩定性檢定：檢驗應變數是否有恆定的統計特性，例如：均值、變異數等，虛無假設：時間序列具有單位根（也就是不穩定），檢定結果 p-value 為 0.0461，拒絕虛無假設，時間序列穩定。

由如圖 5.1 所示，時間序列有持續的自我相關性 (ACF 拖尾)，如圖 5.2 所示，只有前一期與當期有顯著相關 (PACF 於第 2 期截尾)，因此選擇使用一階自我迴歸項 AR(1)，不使用移動平均項 (MA)，如圖 5.3 所示，第三列為季節性檢定，若時間序列具有季節性，則應為單一頻率圖，代表本研究的時間序列並不需要考慮季節性，且時間序列不具單位根，因此本研究不選擇差分，綜上所述，本研究僅選擇模型之參數 AR(1)。

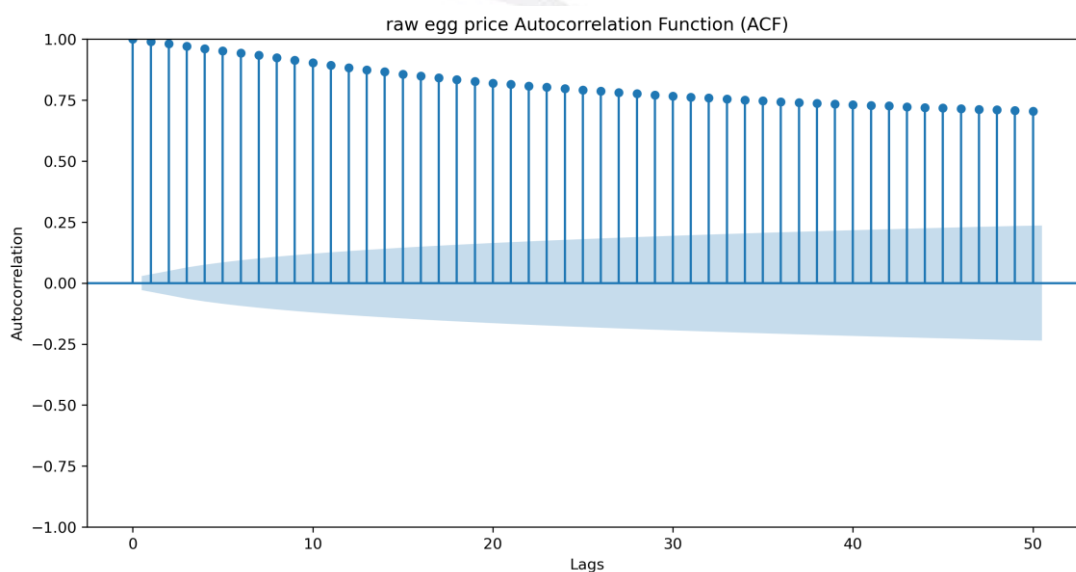


圖 5.1、ACF 圖

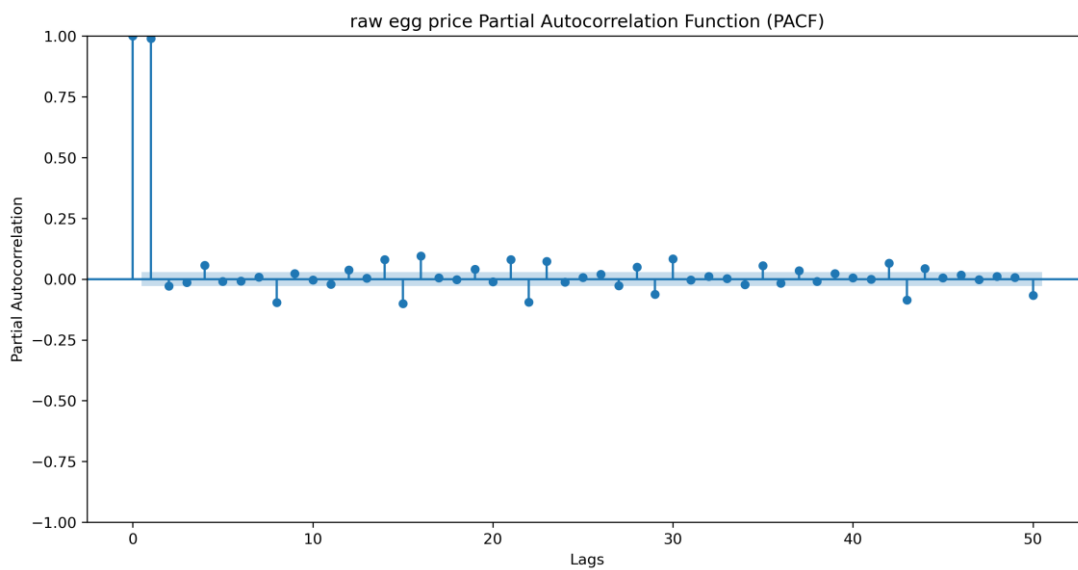


圖 5.2、PACF 圖

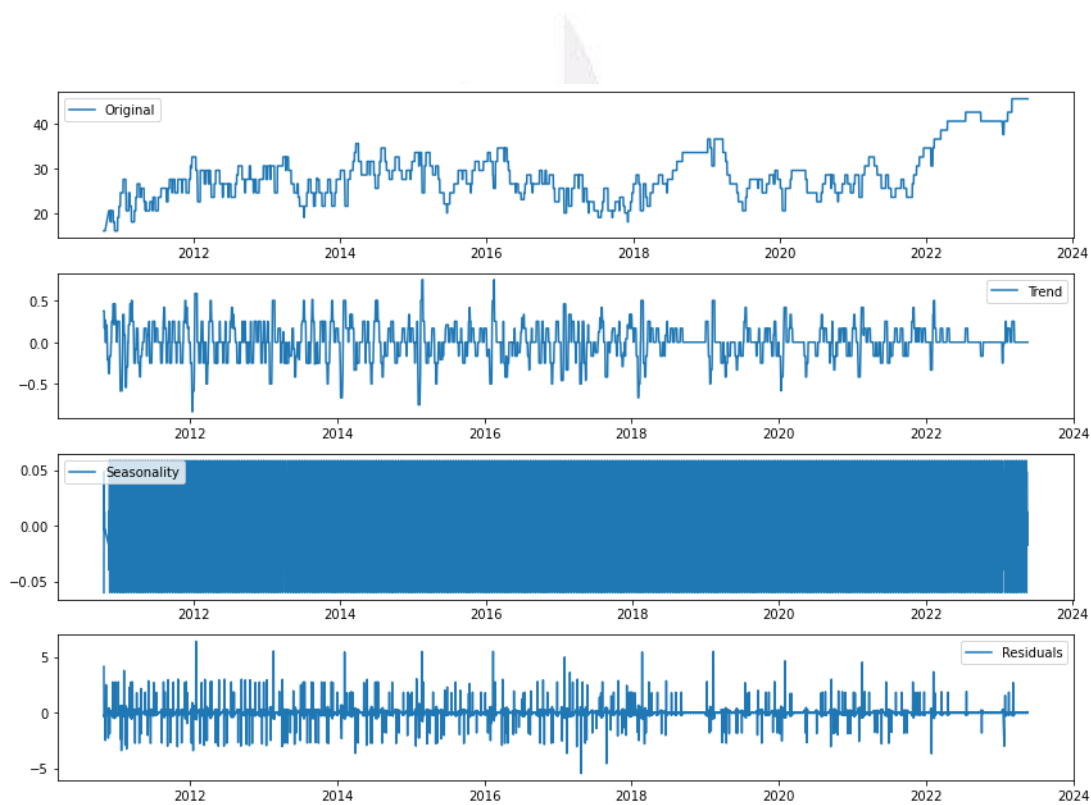


圖 5.3、季節性檢定



ARIMAX 模型程式碼如圖 5.4 所示：

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
from tabulate import tabulate
import matplotlib.pyplot as plt

# 讀取資料
df = pd.read_csv('data/all_data_num_eliminate.csv')

# 提取目標變量
y = df['egg_price']

# 提取原始全部外生變數
X = df[['new_year', 'food_inflation', 'box_egg', 'num_eliminate', 'num_molt', 'feed', 'temp_avg', 'temp_avg_square']]

# 將資料分為訓練集和測試集
train_size = int(len(df) * 0.8)
train_y, test_y = y[:train_size], y[train_size:]
train_X, test_X = X[:train_size], X[train_size:]

# 建立ARIMAX模型
model = sm.tsa.ARIMA(train_y, exog=train_X, order=(1, 0, 0))

# 擬合模型
model_fit = model.fit()

# 預測
predictions = model_fit.predict(start=len(train_y), end=len(train_y)+len(test_y)-1, exog=test_X)

# 評估模型
mse = np.mean((predictions - test_y)**2)
rmse = np.sqrt(mse)

# 計算R平方
y_mean = np.mean(test_y)
total_sum_of_squares = np.sum((test_y - y_mean)**2)
residual_sum_of_squares = np.sum((test_y - predictions)**2)
r_squared = 1 - (residual_sum_of_squares / total_sum_of_squares)
```

圖 5.4、ARIMAX 模型程式碼

$$P_t = c + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varphi_1 P_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$$

上式為 ARIMAX 之方程式， $P_t$  為當期蛋價， $c$  為截距項， $X_1 \sim X_n$  為自變數（外生變數，例如：淘汰隻數）而  $\beta_1 \sim \beta_n$  為自變數之係數， $P_{t-1}$  為前一期的蛋價，而  $\varphi_1$  為其係數， $\epsilon_{t-1}$  為前一期之誤差項， $\theta_1$  為其係數， $\epsilon_t$  為當期誤差項。

模型結果彙整如表 5.1 所示，與多元迴歸模型解釋方法相似，若逢農曆新年，則蛋價會下降 1.6554 元；食物類 CPI 上升一單位，蛋價則上升 0.5016 元。若逢農曆新年、產蛋箱數越多、換羽隻數越多，蛋價會越低；食物類 CPI 越高、前一期氣溫越高、前一期蛋價越高，蛋價則越高，與參考文獻之研究結果相同。 $\sigma^2$  為誤差項之標準差，其顯著代表模型擬合程度不好，無法很好解釋資料的變異性，或有尚未考慮到的因素未放入模型。

表 5.1、ARIMAX 模型結果

	coef	std err	z	P> z
const	11.1949	16.727	0.669	0.503
new_year	-1.6554	0.084	-19.614	0.000
food_inflation	0.5016	0.035	14.276	0.000
box_egg	-0.0003	5.85e-05	-5.629	0.000
num_eliminate	-5.02e-05	4.73e-05	-1.061	0.289
num_molt	-4.144e-05	1.9e-05	-2.185	0.029
feed	-0.1601	0.823	-0.195	0.846
temp_avg	1.2063	0.037	32.456	0.000
temp_avg_square	-0.0332	0.001	-39.345	0.000
ar.L1	0.9602	0.005	192.712	0.000
sigma2	0.8769	0.011	78.587	0.000

如圖 5.5 所示，模型預測效果不理想，RMSE 為 9.3654，本研究認為問題出在模型設定，例如：考慮到運送時間，產蛋箱數可能需要使用前幾期、CPI 為落後指標，應使用前一個月的較為合理。

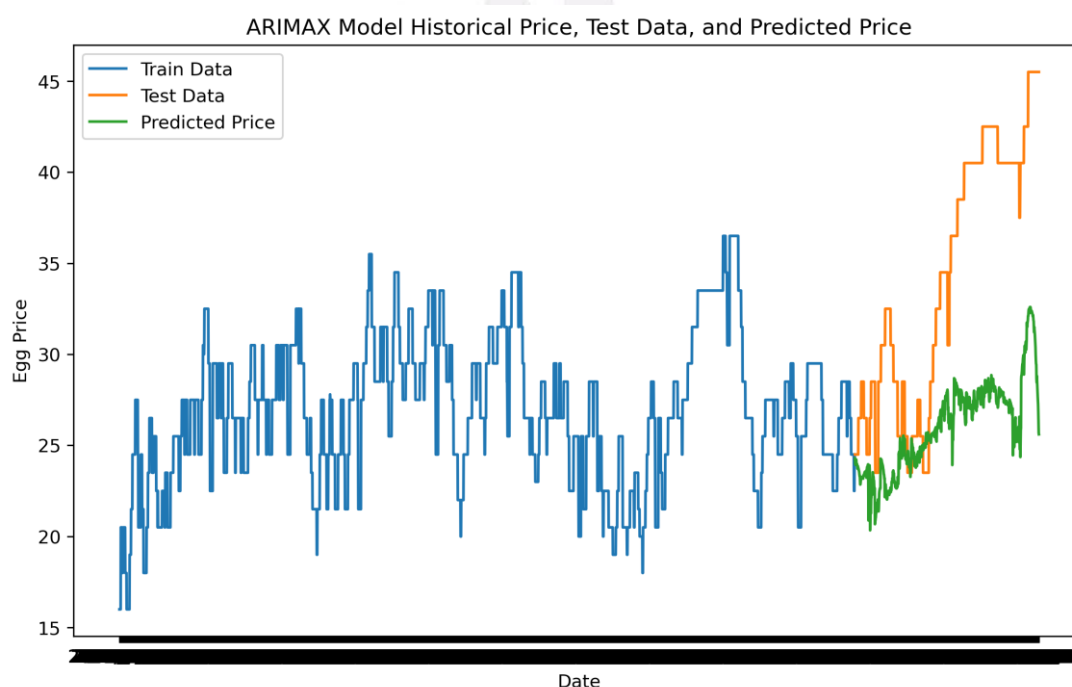


圖 5.5、ARIMAX 模型預測結果

有鑑於食物通膨 (food\_inflation) 和產蛋箱數 (box\_egg) 兩因子可能需要一段時間才會反映在蛋價上，而非直接對當日蛋價產生立即性的影響，因此本研究將食物通膨因子延後 30 天，產蛋箱數因子延後 7 天，也就是透過 30 天後的食物

通膨因子和 7 天後產蛋箱數因子預測當日的蛋價，其餘因子根據唐婉庭 (2019) 使用當天資料預測，以測試是否能提升最終的預測結果，ARIMAX 模型結果如表 5.2 所示，ARIMAX 預測結果如圖 5.6 所示，RMSE 為 13.8329，R 平方為-2.775，與實際價格平均相差大約 14 元，預測結果誤差不減反增。

表 5.2、ARIMAX 模型結果 (食物通膨和產蛋箱數因子延後)

	coef	std err	z	P> z
const	39.8068	20.456	1.946	0.052
new_year	-1.7783	0.085	-20.804	0.000
food_inflation	0.3232	0.039	8.385	0.000
box_egg	-0.0004	7.89e-05	-4.557	0.000
num_eliminate	-1.302e-05	4.81e-05	-0.271	0.787
num_molt	-2.532e-05	1.83e-05	-1.384	0.166
feed	-1.3299	0.988	-1.346	0.178
temp_avg	1.2529	0.031	39.937	0.000
temp_avg_square	-0.0285	0.001	-35.604	0.000
ar.L1	0.9633	0.005	201.730	0.000
sigma2	0.8381	0.010	83.146	0.000

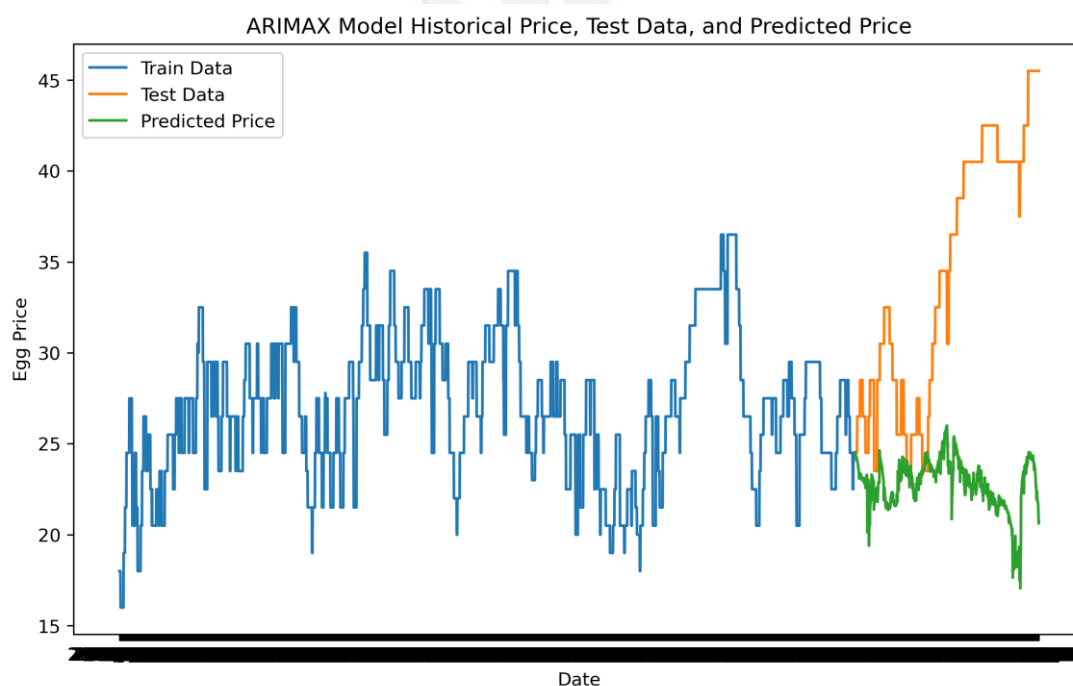


圖 5.6、ARIMAX 模型預測結果 (食物通膨和產蛋箱數因子延後)

## 六、 Long Short-Term Memory (LSTM) Model

### 1. LSTM 模型架構

LSTM 為類神經網路模型，如圖 6.1 所示包含輸入層、隱藏層和輸出層，以本研究的蛋價預測模型為例，本研究之 LSTM 模型輸入數為 8 輸出數為 1，表示透過 8 個特徵預測 1 個價格。

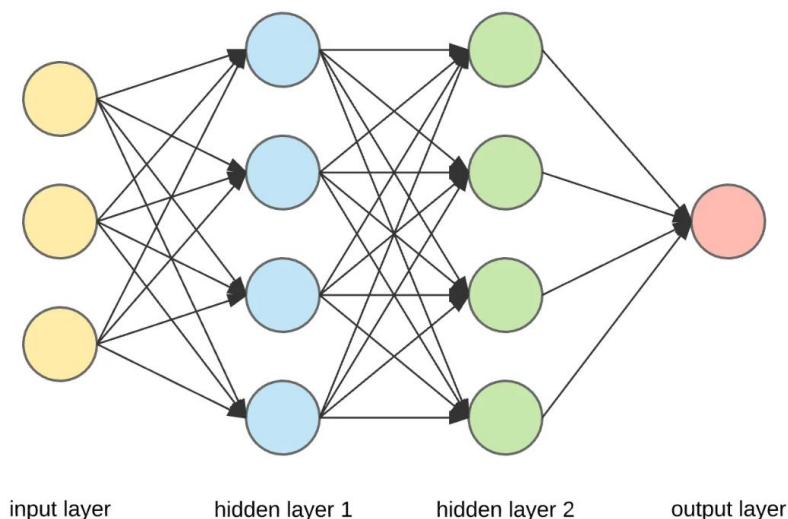
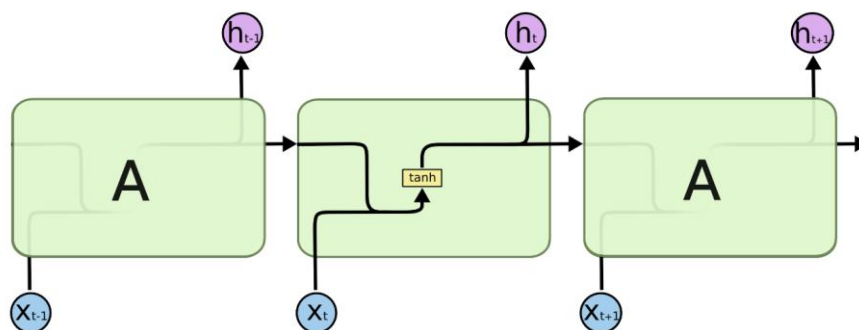


圖 6.1、類神經網路架構

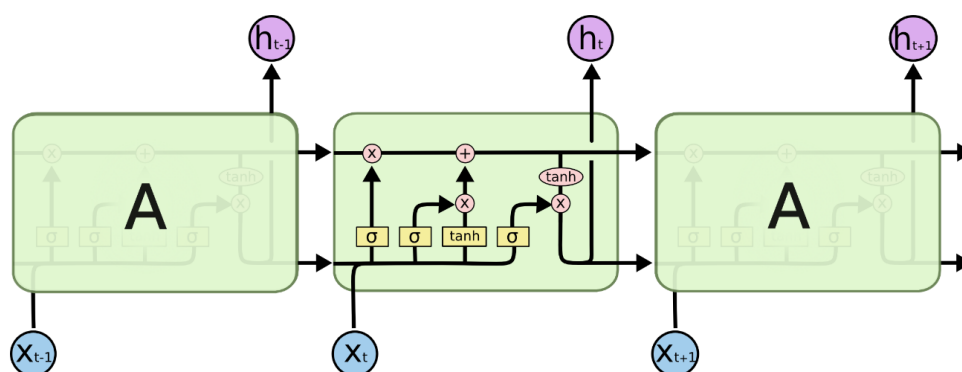
LSTM 為循環神經網路 (Recurrent Neural Network, RNN) 之延伸，RNN 的模型架構 (如圖 6.2 所示) 特色為會將上一個輸出加入至下一個的輸入，使模型具有歷史記憶性，但此架構較為單純，容易使較久但仍然重要影響當期資料的歷史資訊，經過多次迭代後，權重低至幾乎無法影響當期的輸出。



The repeating module in a standard RNN contains a single layer.

圖 6.2、RNN 模型架構 (Christopher Olah, 2015)

LSTM 模型架構如圖 6.3 所示，以 RNN 為基礎加入 Input Gate、Output Gate 和 Forget Gate 分別控制輸入、輸出和遺忘的權重，該權重由激活函數 Sigmoid 轉換為 0 至 1 的數值（如圖 6.4 所示），並且將輸入透過激活函數 tanh 轉換為-1 至 1 的數值（如圖 6.5 所示），相當於將輸入由 1 個變為 4 個，權重透過梯度下降和反向傳播法讓模型自主學習，讓模型具有遺忘和記憶的能力，以提升模型的預測效果。



The repeating module in an LSTM contains four interacting layers.

圖 6.3、LSTM 模型架構 (Christopher Olah, 2015)

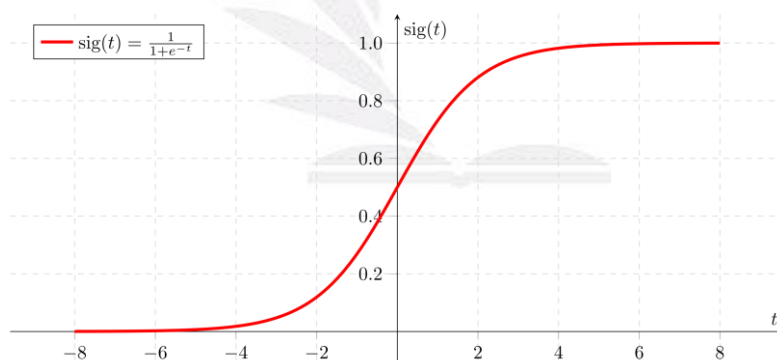


圖 6.4、Sigmoid 激活函數

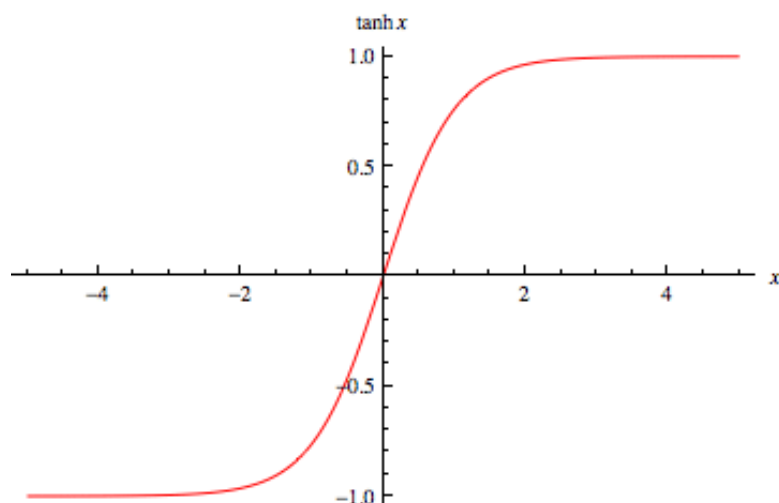


圖 6.5、Tanh 激活函數

## 2. LSTM 模型訓練與參數調整

本研究採用雙層的 LSTM 進行訓練和預測，首先將資料進行正規化後，將 80% 的資料輸入至模型中進行訓練，Epochs 為 100 訓練結果之 Loss 折線圖如圖 3.6 所示，訓練結果為欠擬合，預測結果如圖 3.7 所示，預測效果很差不可用。



圖 6.6、初始 LSTM 訓練結果

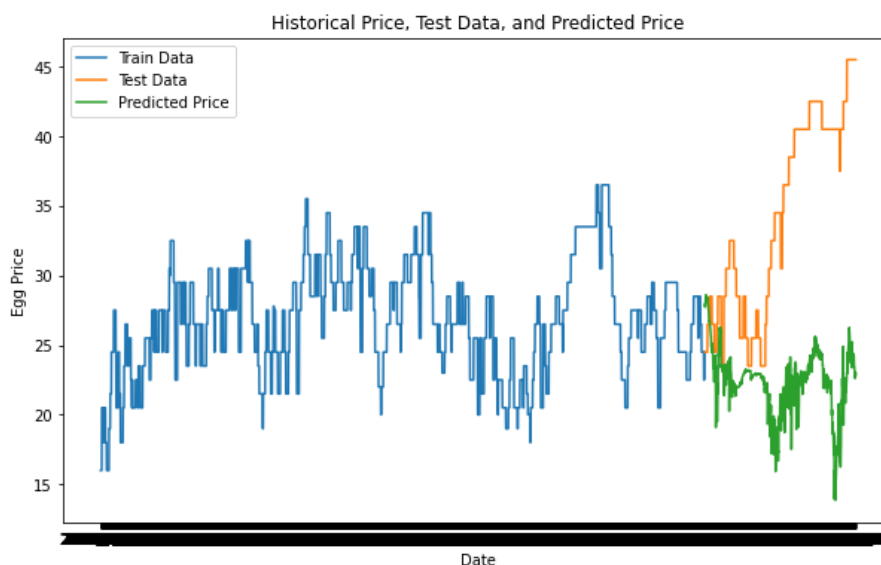


圖 6.7、初始 LSTM 預測結果

推測有可能為資料量太少，因此模型過早學習到不必要的特徵，因此在訓練集 Loss 下降過早，但在測試集 Loss 無法下降判定為訓練失敗，因此本研究透過 Sliding Window 將時序資料變成許多筆短短的時序資料，重複每間隔一天就取一周的資料，讓資料集數量擴充為 7 倍，Stride (步伐) 相當於 1，window-size (窗口大小) 相當於 7，等同於每隔一天就取一個星期的資料，並且在隱藏層中間加入 20% 的 Dropout，訓練結果如圖 3.8 所示，發現模型產生過擬合，測試集在 Epochs 為 15 至 20 左右降至最低，而後逐漸上升，表示模型訓練過頭。

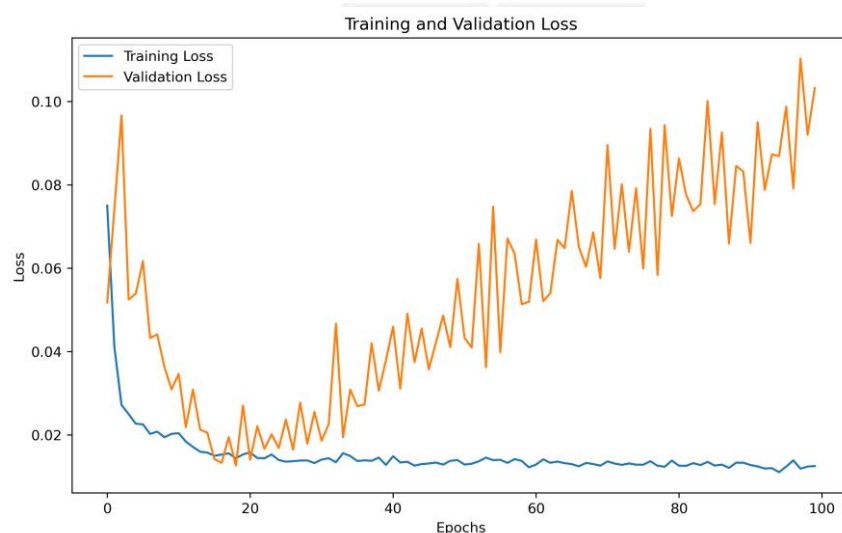


圖 6.8、LSTM 過擬合訓練結果

因此本研究將 Epochs 調整為 15，訓練結果如圖 3.9 所示，模型訓練成功，最終模型預測結果如圖 3.10 所示，RMSE 為 3.0985，LSTM 資料前處理程式碼如圖 3.11 所示，最終模型程式碼如圖 3.12 所示，模型參數如表 3.1 所示。

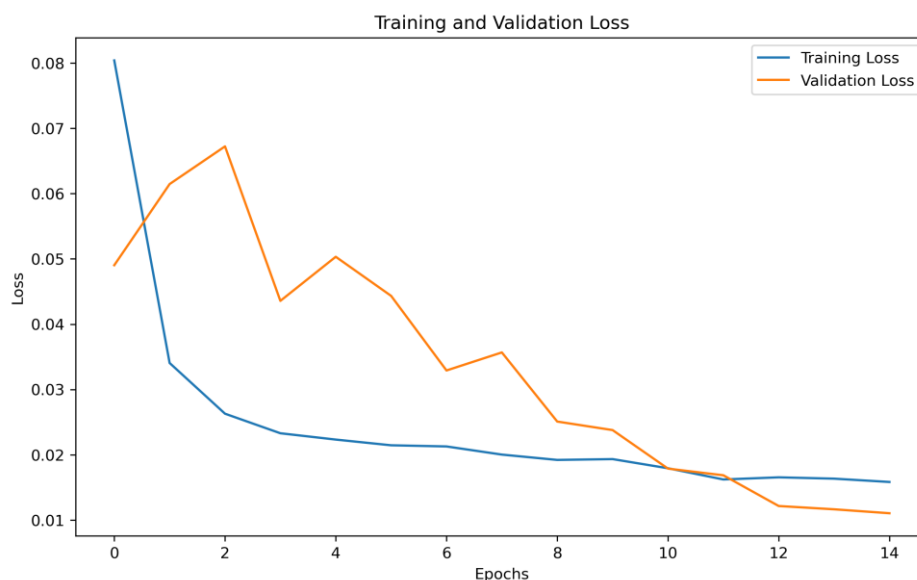


圖 6.9、LSTM 訓練成功結果

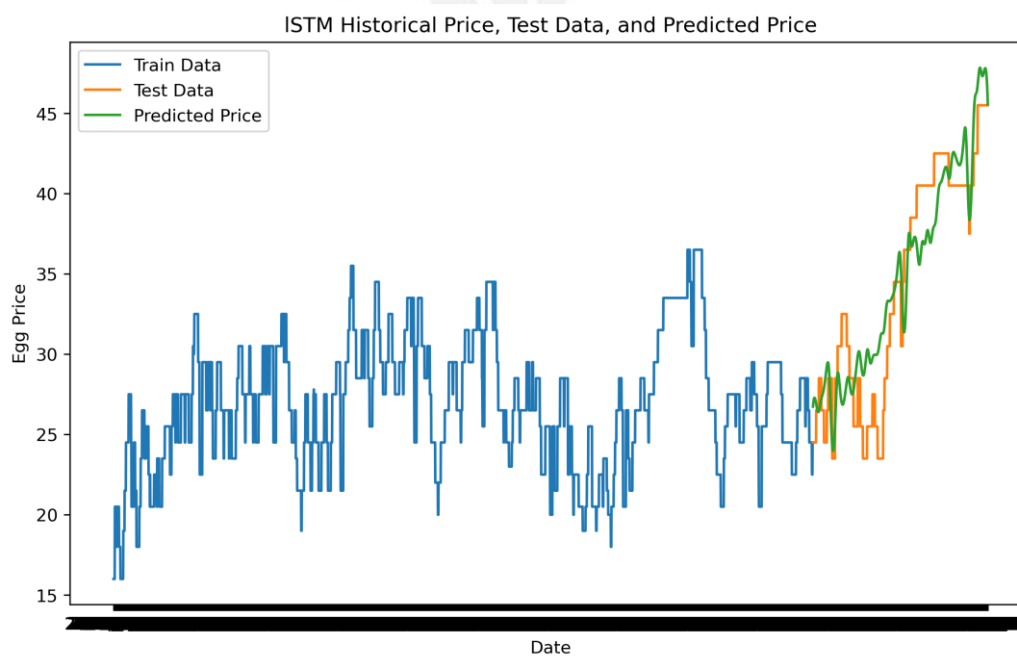


圖 6.10、LSTM 最終預測結果



```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('data/all_data_num_eliminate.csv')

# 將日期欄位轉換為datetime格式並設定為索引
df['date'] = pd.to_datetime(df['date'])
df.set_index('date', inplace=True)

# 擷取目標變量
y = df['egg_price']

# 擷取外生變數
X = df[['new_year', 'full_inflation', 'food_inflation', 'num_chicken', 'box_egg', 'num_eliminate', 'num_molt', 'feed', 'corn', 'temp_avg', 'temp_avg_square']]

# 將資料正規化
scaler_y = MinMaxScaler()
scaled_y = scaler_y.fit_transform(y.values.reshape(-1, 1))

scaler_X = MinMaxScaler()
scaled_X = scaler_X.fit_transform(X)

def create_time_series_data(X, y, time_steps, window_size):
    X_time_series = []
    y_time_series = []
    for i in range(0, len(X) - time_steps, window_size):
        X_window = X[i:i+time_steps]
        y_window = y[i:i+time_steps-1]
        X_time_series.append(X_window)
        y_time_series.append(y_window)
    return np.array(X_time_series), np.array(y_time_series)
```

圖 6.11、LSTM 資料前處理

```
# 設定時間步長 (Time Steps) 和滑動窗口大小 (Window Size)
time_steps = 1
window_size = 7

# 將目標變量也正規化
scaled_y = scaler.fit_transform(y.values.reshape(-1, 1))

# 轉換資料為時間序列數據
X_time_series, y_time_series = create_time_series_data(scaled_X, scaled_y, time_steps, window_size)

# 計算測試集的大小
test_size = int(len(X_time_series) * 0.2)

# 分割訓練集和測試集
train_X, test_X = X_time_series[:-test_size], X_time_series[-test_size:]
train_y, test_y = y_time_series[:-test_size], y_time_series[-test_size:]

# 建立LSTM模型
model = Sequential()
model.add(LSTM(64, input_shape=(time_steps, X.shape[1]), return_sequences=True))
model.add(Dropout(0.2)) # 添加 Dropout 層
model.add(LSTM(64))
model.add(Dropout(0.2)) # 添加 Dropout 層
model.add(Dense(1))
model.compile(optimizer='adam', loss='mean_squared_error')

# 訓練模型
history = model.fit(train_X, train_y, epochs=15, batch_size=32, verbose=2, validation_data=(test_X, test_y))

# 預測
predictions = model.predict(test_X)

# 將預測結果轉換為原始尺度
predicted_values = scaler.inverse_transform(predictions)

# 評估模型
mse = np.mean((predicted_values - scaler.inverse_transform(test_y.reshape(-1, 1)))**2)
rmse = np.sqrt(mse)
```

圖 6.12、LSTM 模型程式碼

表 6.1、LSTM 模型參數

參數名稱	參數
Cells	64
Layers	2
Dropout	0.2
optimizer	adam
loss	mean_squared_error
epochs	15
batch_size	32



## 七、 結論與未來展望

### 1. 實用性

本研究透過迴歸分析和類神經網路方式，取代人工預測蛋價的，並改為從大數據中找到可能影響蛋價的因子和趨勢，以預測未來有可能的漲價期間，從而提早制定措施預防避免民眾恐慌，並且有利於政府或民間相關單位政策制定。

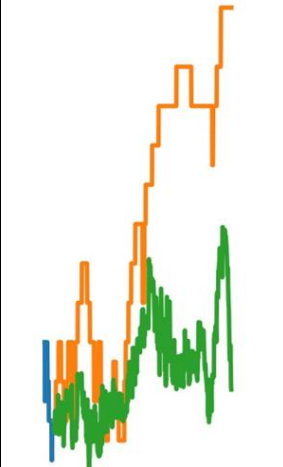
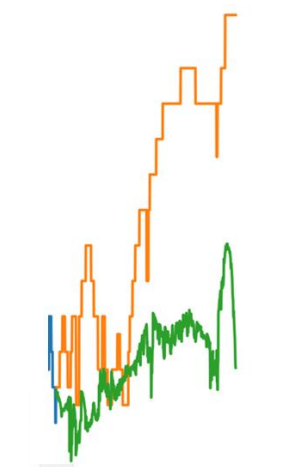
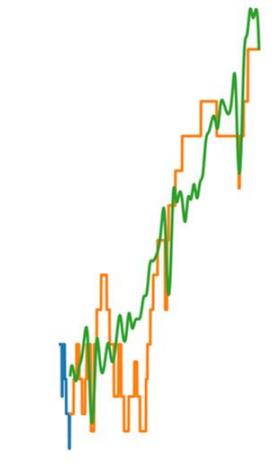
### 2. 創新性

本研究以時間序類分析的類神經架構 LSTM 對蛋價進行預測，有別於過往採用機器學習如多元迴歸分析等方法，需進行多重的特徵工程步驟，確保各項特徵對於預測價格有效、正確且不會互相干擾，LSTM 僅需制定好架構，並且合理的調整參數，使模型預測效果達到最佳，而不用對特徵進行較多的處理。

### 3. 貢獻與效益

如表 7.1 所示，本研究最終的研究成果為使用 LSTM 的預測效果最好，過往較無使用類神經網路，針對臺灣蛋價進行預測，因此本研究提供一個有效且可行的方式，並於本文中說明 LSTM 的架構與參數，可供後續研究人員或是政府單位建立模型之參考。

表 7.1、各模型預測結果彙整

模型名稱	多元迴歸模型	ARIMAX	LSTM
RMSE	9.2412	9.3654	3.0985
預測時間	2020/10/29 至 2023/04/30		
預測結果圖 橫軸：時間 縱軸：蛋價			

#### 4. 未來展望

在 LSTM 模型預測效果為三個模型之中最佳，但 3 塊錢的誤差仍較大，可能參數未調整至最佳，可以再透過參考更多研究成果以調整神經網路架構，或是透過反覆調整參數進行訓練，以精進 LSTM 的預測效果。此外，深度學習為黑盒模型，相較於時間序列分析的可解釋性較差，因此仍須精進時間序列分析模型之預測準確度，供政府單位參考，才能進一步透過判斷各項因子對於蛋價的影響，使政策能夠更加對症下藥，精準地解決蛋價浮動的問題。

多元迴歸模型與 ARIMAX 模型之表現不甚理想，原因除了變數的選擇，還有模型設定上的問題，變數的共線性問題可以使用變異數膨脹因子 (VIF) 檢測，時間序列模型在變數調整的需考量更多，例如：產蛋箱數、CPI 等受時間影響的變數。使用日資料需更小心處理變數、模型，才能發揮模型的預測能力。此外，本研究透過將食物通膨因子延後 30 天，產蛋箱數因子延後 7 天，在多元迴歸模型與 ARIMAX 模型表現皆不盡理想，甚至更差，本研究在延後因子時間手法上過於粗糙，僅透過作者之主觀判斷，應透過詢問專業人員意見，以評估該手法之正確性和有效性，才能真正提升模型預測表現。

接著，本研究尚未建立網頁或是應用程式，並無提供使用者介面方便使用者查看與互動較為可惜，本研究預測蛋價與民眾的生活息息相關，期望未來能模仿

如天氣預報等應用程式，建立一個介面開放給一般大眾使用，提升本研究建構模型之效益。

最後，於資料方面希望政府將禽流感撲殺數（並無官方彙整的資料集）、淘汰蛋雞數量、換羽蛋雞數量等與蛋價相關資料收集的更加完善，並將資料集以檔案、網頁形式或提供 API 以便透過爬蟲取得資料，形成動態資料，透過不斷自動更新資料，使模型更加準確，也能提供給使用者及時資訊。



## 參考資料

1. 王柏鈞，2020，「時間序列的 AI 介紹」，Medium 文章
2. 唐婉庭，2019，「以時間序列法檢測臺灣雞蛋產地價格模型與預測比較分析」，國立中興大學生物產業管理研究所碩士學位論文
3. Christopher Olah, 2015, “Understanding LSTM Networks.”
4. Qingqi Zhang, “Housing Price Prediction Based on Multiple Linear Regression”, *Scientific Programming*, vol. 2021, Article ID 7678931, 9 pages, 2021. <https://doi.org/10.1155/2021/7678931>

