

國立臺灣大學社會科學院經濟學系



碩士論文

Department of Economics

College of Social Sciences

National Taiwan University

Master Thesis

匯率預測：

結合時間序列模型與自注意力機制

Forex Prediction:

Combining Time Series Model and Self-Attention  
Mechanism

簡崇益

Chung-I Chien

指導教授: 林建甫 博士

Advisor: Chien-Fu Lin Ph.D.

中華民國 111 年 4 月

April, 2022

# 國立臺灣大學碩士學位論文

## 口試委員會審定書



匯率預測：

結合時間序列模型與自注意力機制

Forex Prediction:

Combining Time Series Model and Self-Attention  
Mechanism

本論文係簡崇益君 (R07323014) 在國立臺灣大學經濟學系完成之碩士學位論文，於民國 111 年 4 月 15 日承下列考試委員審查通過及口試及格，特此證明

口試委員： 林建甫

(指導教授)

許振明

謝德宗

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

王柏元

\_\_\_\_\_

\_\_\_\_\_

所長： \_\_\_\_\_





## 摘要

匯率預測一直以來都是學界、業界和政府單位關心的議題，匯率的變動可能會影響實體經濟活動和金融市場。過去許多研究顯示，傳統的經濟模型在匯率的樣本外預測方面無法打敗隨機漫步模型，因為傳統模型大多為線性模型，而匯率序列本質上是動態的和非線性的。近年來有許多深度模型被提出，雖然其主要是用在自然語言處理的研究上，但其在捕捉非線性特性的卓越能力，使得越來越多研究開始嘗試將其應用在匯率預測方面。

本文主要參考 Yilmaz and Arabaci (2021) 的模型，將匯率的預測拆分成線性和非線性兩個部分，分別使用整合移動平均自迴歸 (Autoregressive Integrated Moving Average, ARIMA) 模型和自注意力 (Self-Attention, SA) 機制分別對美元兌加元及澳幣、英鎊兌美元匯率之報酬率進行預測，將此模型與 Yilmaz and Arabaci (2021) 研究中使用長短期記憶 (Long Short-term Memory, LSTM) 架構的 ARIMA-LSTM 模型及隨機漫步模型進行比較後，結果顯示改用自注意力機制的 ARIMA-SA 模型的預測能力較 ARIMA-LSTM 模型來得差，自注意力機制在匯率預測方面無法得到好於長短期記憶機制的結果。甚至，不同於 Yilmaz and Arabaci (2021) 所顯示的結果，ARIMA-LSTM 模型在日匯率報酬方面的預測能力亦劣於隨機漫步模型。

關鍵字：匯率、預測、時間序列、深度學習、混合模型





# Abstract

Exchange rate forecasting has long been an issue of interest to academics, industries, and governments, where changes in exchange rates may affect real economic activity and financial markets. Many studies have shown that traditional economic models cannot beat random walk models for out-of-sample forecasting of exchange rates, because most of the traditional models are linear, while exchange rate series are inherently dynamic and non-linear. In recent years, many deep models have been proposed, and although they are mainly used in natural language processing studies, their excellent ability to capture nonlinear properties has led more and more studies to try to apply them in exchange rate prediction.

This paper mainly refers to the model from Yilmaz and Arabaci (2021), which splits the forecast of exchange rate into linear and non-linear components, using the Autoregressive Integrated Moving Average (ARIMA) model and Self-Attention (SA) mechanism to forecast the return of USD/CAD, AUD/USD, and GBP/USD respectively. Comparing

this model with the ARIMA-LSTM model using the Long Short-term Memory (LSTM) framework in Yilmaz and Arabaci (2021) and the random walk model, the results show that the ARIMA-SA model has worse predictive power than the ARIMA-LSTM model.

Moreover, unlike the results shown by Yilmaz and Arabaci (2021), the predictive power of the ARIMA-LSTM model in terms of daily exchange rate returns is inferior to that of the random walk model.

**Keywords:** Forex, Prediction, Time Series, Deep Learning, Hybrid model



# 目錄

	<b>Page</b>
口試委員審定書	i
摘要	iii
<b>Abstract</b>	<b>v</b>
目錄	<b>vii</b>
圖目錄	<b>ix</b>
表目錄	<b>x</b>
第一章 緒論	<b>1</b>
1.1 研究動機	1
1.2 研究架構	2
1.3 研究流程	2
第二章 文獻回顧	<b>4</b>
2.1 傳統經濟計量模型相關文獻	4
2.2 類神經網路模型相關文獻	5
2.3 混合模型相關文獻	6
第三章 實證模型	<b>7</b>
3.1 線性部分	7
3.1.1 ARIMA	7

3.2	非線性部分	8
3.2.1	長短期記憶 (LSTM)	8
3.2.2	自注意力 (SA)	10
<b>第四章</b>	<b>實證方法與結果</b>	<b>13</b>
4.1	資料說明及資料集分割	13
4.2	研究方法	14
4.3	實證結果	20
<b>第五章</b>	<b>結論</b>	<b>22</b>
5.1	研究結論	22
5.2	研究建議	23
	參考文獻	24





## 圖目錄

1.1	研究流程圖 . . . . .	3
3.1	模型建構示意圖 . . . . .	7
3.2	LSTM 示意圖 . . . . .	9
3.3	Self-Attention 示意圖-1 . . . . .	10
3.4	Self-Attention 示意圖-2 . . . . .	11
3.5	Self-Attention 示意圖-3 . . . . .	11
3.6	Self-Attention 示意圖-4 . . . . .	11
4.1	美元兌加元匯率 . . . . .	15
4.2	澳幣兌美元匯率 . . . . .	15
4.3	英鎊兌美元匯率 . . . . .	16
4.4	美元兌加元匯率取對數後一階差分 . . . . .	16
4.5	澳幣兌美元匯率取對數後一階差分 . . . . .	17
4.6	英鎊兌美元匯率取對數後一階差分 . . . . .	17
4.7	滾動與遞迴視窗示意圖 . . . . .	18



## 表目錄

4.1	變數說明 . . . . .	13
4.2	變數敘述統計 . . . . .	14
4.3	超參數網格搜索範圍 . . . . .	19
4.4	各模型匯率預測結果 . . . . .	20
4.5	各模型殘差分析 . . . . .	21



# 第一章 緒論

## 1.1 研究動機

外匯匯率是所有經濟體的重要金融和總體經濟變數之一，匯率變動不僅會顯著影響對外貿易、資本流動和資產組合管理，還會顯著影響實體經濟活動。顯然，考慮到它們產生的風險，匯率變動增加了商業和投資決策的難度。因此，匯率變動的預測一直受到學術界、政府單位和企業的關心。

從交易量來看，根據國際清算銀行 (BIS, Bank for International Settlements) 的統計，外匯市場平均每天交易量超過六兆美元，相較之下股票市場的日交易量只有約四千億美元，這表示外匯市場不僅是最大的金融交易市場，其市場規模更遠超股票市場，外匯市場的風吹草動亦時刻牽動金融機構及投資人的決策判斷。

過去在國際金融領域已經發展了許多經濟模型試圖去解釋匯率的變動，但是在實證上，這些模型卻常常無法打敗簡單的隨機漫步模型 (Random Walk Model)。準確地說，當使用浮動匯率的資料去測試這些經濟模型時，所得到的樣本外預測結果並不優於隨機漫步。

隨著科技的演進，電腦運算能力快速提升，讓類神經網路以各式不同的架構應用於廣泛領域，如卷積神經網路 (Convolutional Neural Networks, CNN) 在影像辨識方面、循環神經網路 (Recurrent Neural Network, RNN) 和後來以自注意力

(Self-Attention) 機制為主建立的 Transformer 模型在語言翻譯與聲音辨識等自然語言處理領域的優異表現。許多文獻與研究機構都指出，類神經網路模型能夠有效找出模型輸入與輸出之間的關聯，從而得到準確的預測結果。



由於類神經網路模型在語言翻譯與聲音辨識等自然語言處理領域的表現非常出色，許多學者也將其應用在金融時間序列的預測上，最近的研究顯示，結合傳統的時間序列模型以及前述的循環神經網路模型，能在月匯率預測上得到很好的結果。而本研究試圖進一步探討，若是將月資料換成日資料，將循環神經網路模型改成以自注意力為主的神經網路模型，在不同的時間頻率及模型架構下，還不能得到一樣良好的結果。

## 1.2 研究架構

本研究共有五個章節。第一章為緒論，說明研究動機及架構。第二章為文獻回顧，介紹此前研究的進展。第三章將會簡單介紹本研究所使用的實證模型。第四章為實證方法與結果，將詳述實證方法，包含所使用資料的說明、資料集如何分割，以及研究方法的闡述，並針對模型的樣本外表現進行實證結果的比較。最後，第五章將為本研究進行結論。

## 1.3 研究流程

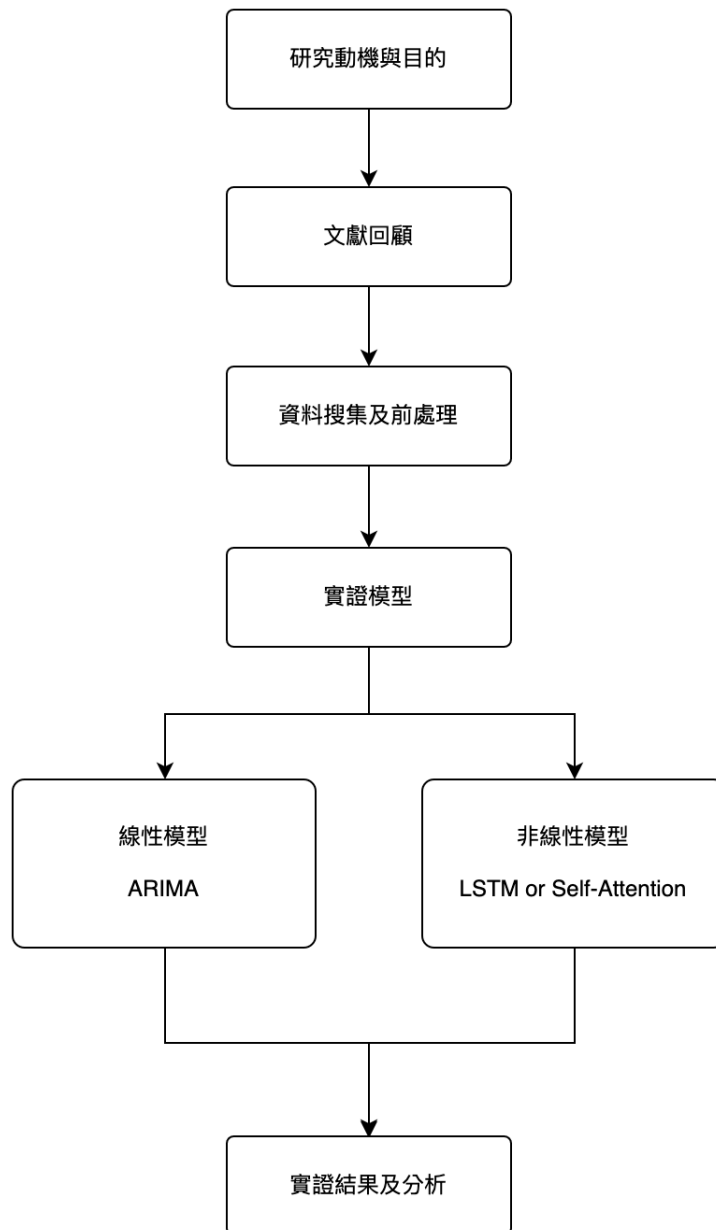


Figure 1.1: 研究流程圖



## 第二章 文獻回顧

自 90 年代以來，許多研究都將興趣集中在建模和預測匯率上，以下將回顧過去相關研究文獻。

### 2.1 傳統經濟計量模型相關文獻

Meese and Rogoff (1983) 使用 1976 至 1981 年美元對英鎊、馬克及日圓匯率，比較了各種結構和時間序列匯率模型，並未發現擁有好的樣本外 (Out-of-sample) 預測能力的模型，即使使用的變數包含未來解釋變數的實現值。Alexander and Thomas III (1987) 拓展上述研究至 1985 年，且包含使用時變參數 (Time-varying Parameters) 的模型，結果顯示在預測範圍 (Forecasting Horizon) 小於兩年時，與 Meese and Rogoff (1983) 有相同結論。Wolff (1988) 則證明 Alexander and Thomas III (1987) 有計算錯誤，並將其結論更正為在所有預測範圍內都與 Meese and Rogoff (1983) 一致。

Rogoff and Stavrakeva (2008) 檢驗了後續文獻發展中預測表現良好的結構模型在不同樣本外檢定及預測視窗 (Forecast Window) 下的穩固性 (Robustness)，認為其結果是不穩定且被誇大的。Rossi (2013) 亦針對匯率預測相關文獻進行批判性回顧，指出儘管部分模型可以成功預測匯率，但這些成功取決於變數、預測範圍、樣本期間和評價方法的選擇，並沒有系統性的實證證據支持匯率的可預測性。因

此，學界的主流看法認為匯率變動仍無法被預測。

線性模型的限制是相關文獻中主要討論匯率不可預測性時最主要的問題。Alvarez-Diaz (2008) 使用 KNN (K-nearest Neighbour) 的方法來預測日圓、英鎊兌美元匯率，認為相對隨機漫步模型來說不好的樣本外表現可能來自於此限制。Engle and Hamilton (1990) 使用隨機分段趨勢模型 (Model of Stochastic Segmented Trends) 捕捉美元匯率的長擺 (Long Swing) 現象，證明該非參數化 (Non-parametric) 模型表現優於隨機漫步模型。

## 2.2 類神經網路模型相關文獻

Hornik et al. (1989) 證明類神經網路模型 (Neural Network, NN) 是通用逼近器 (Universal Approximator)，只要經過足夠的學習，擁有足夠的隱藏神經元 (Neuron)，就能夠成功逼近任意函數。隨著電腦運算能力快速提升，類神經網路以各式不同的架構成功應用於廣泛領域，如卷積神經網路在影像辨識方面、循環神經網路和後來以自注意力機制為主建立的 Transformer 模型在語言翻譯與聲音辨識等自然語言處理領域的發展，也因此匯率預測相關文獻對這類模型的興趣漸增。

根據 Yilmaz and Arabaci (2021)，早在 1992 年就出現了第一個使用類神經網路預測匯率的案例，其結果顯示類神經網路的預測能力好於隨機漫步模型。Tenti (1996) 和 Giles et al. (2001) 則是使用循環神經網路達到很好的預測表現，而根據 Kiani and Kastens (2008) 對一般類神經網路、循環神經網路和線性模型的比較，英鎊和日圓都分別有循環神經網路表現優於線性模型，但加元則沒有神經網路的表現優於線性模型。

在金融時間序列預測中，循環神經網路的變形—長短期記憶 (Long Short-term

Memory, LSTM) 常常是研究者偏好的選擇 (Sezer et al., 2020; Wu and Gao, 2018), 其改善了傳統循環神經網路在訓練過程中容易出現的梯度消失 (Gradient Vanishing) 的問題, 也在許多研究中表現好於一般循環神經網路 (Qu and Zhao, 2019; Ranjit et al., 2018; Dodevski et al., 2018)。

## 2.3 混合模型相關文獻

有一些研究者嘗試使用混合模型 (Hybrid Model) 和集成方法 (Ensemble Method) 來改善預測表現, Lai et al. (2006) 表示混合和集成方法常常在匯率預測方面得到相較單一模型要好的結果。Yilmaz and Arabaci (2021) 則透過將模型拆分成線性和非線性部分, 結合 ARIMA 和 LSTM, 在加元、澳幣、英鎊月匯率預測方面顯著優於其他模型。

本文主要參考 Yilmaz and Arabaci (2021) 來建構美元兌加元及澳幣、英鎊兌美元日匯率的預測模型, 將匯率的預測拆分成線性和非線性兩個部分, 其中線性的部分一樣採用 ARIMA 模型, 而非線性的部分則使用可以平行化運算的自注意力機制。之所以採用自注意力機制, 是考慮到從月資料換成日資料, 匯率預測所需的序列長度可能較長, 而一般循環神經網路或 LSTM 對於比較久遠的輸入比較不能有效學習, 以及為了探究平行化運算帶來的速度上的好處會不會伴隨著預測表現的下降, 抑或是反而表現得更好。



## 第三章 實證模型

本文依循 Yilmaz and Arabaci (2021) 的方式建構模型 (圖 3.1)。<sup>1</sup> 將匯率的預測拆分成線性和非線性兩個部分，先使用 ARIMA 模型預測線性的部分，再分別使用原文所用的 LSTM 及本文想探討的自注意力機制來預測前述線性模型的殘差，形成最終預測。以下簡要介紹 ARIMA、LSTM 和 SA。

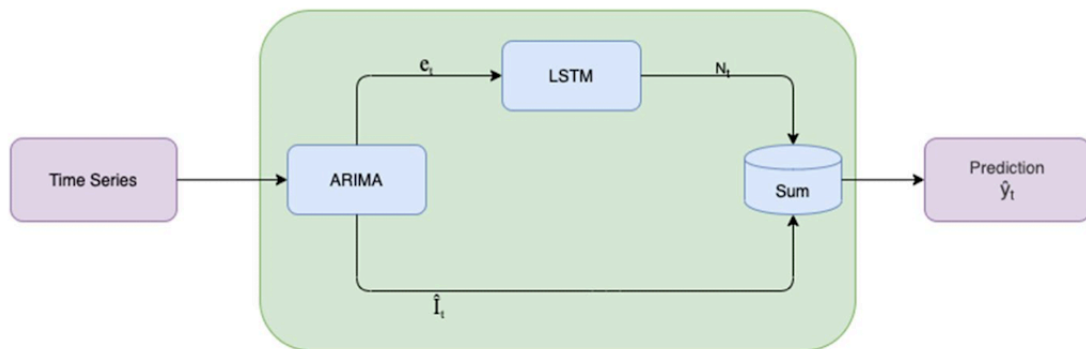


Figure 3.1: 模型建構示意圖

### 3.1 線性部分

#### 3.1.1 ARIMA

整合移動平均自迴歸模型 (Autoregressive Integrated Moving Average model, ARIMA) 是研究時間序列的重要方法，透過適當的差分，結合自迴歸模型

<sup>1</sup>圖片來源：Yilmaz and Arabaci (2021)

(Autoregressive model, AR) 與移動平均模型 (Moving Average model, MA) 來進行預測，通常寫成 ARIMA(p,d,q)，其形式如下：



$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - L)^d y_t = (1 + \sum_{i=1}^q \theta_i L^i) \epsilon_t \quad (3.1)$$

其中  $L, y_t, \epsilon_t$  分別代表落後運算子，第  $t$  期變數和第  $t$  期的隨機項。

## 3.2 非線性部分

### 3.2.1 長短期記憶 (LSTM)

長短期記憶 (Long Short-term Memory, LSTM) 首先由 Hochreiter and Schmidhuber (1997) 所提出，是一種特殊的循環神經網路 (Recurrent Neural Network, RNN)，主要是用來解決一般循環神經網路在訓練時容易出現的梯度消失問題，改善一般循環神經網路對較久遠的輸入，訓練不效率的狀況。

LSTM 的神經元主要由三個閘 (Gate) 所構成，分別是：輸入閘 (Input Gate)、遺忘閘 (Forget Gate)、輸出閘 (Output Gate)。

架構如圖 3.2 所示，<sup>2</sup>其中  $x_t, c_t, h_t$  分別表示在時間為  $t$  的輸入向量、神經元狀態及神經元的輸出。若以數學式表達則為：<sup>3</sup>

(輸入閘：決定經過轉換的當期輸入變量  $h_{t-1}, x_t$  有多少比例要保留在當期神經元中)

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i)$$

<sup>2</sup>圖片來源：Yilmaz and Arabaci (2021)

<sup>3</sup>其中  $\sigma(z) = \frac{1}{1+e^{-z}}$ ， $\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ ， $\odot$ ：相同大小的向量各元素相乘。

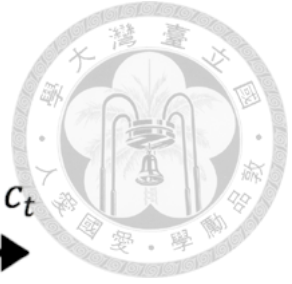
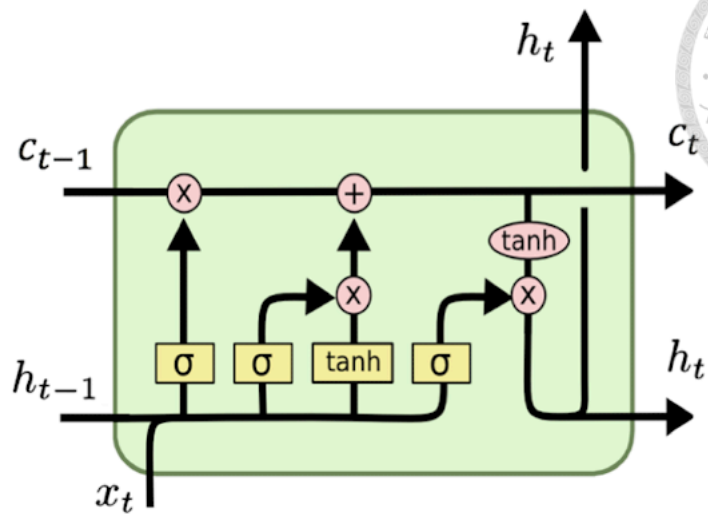


Figure 3.2: LSTM 示意圖

(輸出閘：決定經過轉換的當期神經元狀態  $c_t$  有多少比例要延續到下期神經元中)

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o)$$

(遺忘閘：決定前一期所輸出的神經元狀態  $c_{t-1}$  有多少比例保留在當期神經元中)

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f)$$

(神經元狀態)

$$c_t = f_t \odot c_{t-1} + i_t \odot \phi(W_c h_{t-1} + U_c x_t + b_c)$$

(神經元輸出)

$$h_t = o_t \odot \phi(c_t)$$

其中  $x_t, h_t$  為 LSTM 在時間點  $t$  的輸入與輸出， $W_k, U_k, b_k, k = i, o, f$  為被訓練的參數。



### 3.2.2 自注意力 (SA)

在 Transformer 架構被 Google Vaswani et al. (2017) 提出以前，自然語言處理包含情感分析、機器翻譯、語音辨識等時間序列相關深度學習任務的 SOTA (State Of The Art) 方法，通常以循環神經網路中的 LSTM 和 GRU (Gated Recurrent Unit, LSTM 的簡化版本) 為主，但其缺點為序列架構，必須計算完前一個時間點的結果，才能進行後續的計算，無法平行化處理以充分利用 GPU 的運算能力，因此模型的訓練速度較慢。在 Transformer 架構中，Google 提出了一個完全避免使用循環神經網路的自注意力 (Self-Attention, SA) 機制，該架構不僅大幅提升訓練速度，更突破以往 SOTA 模型的表現，成為後來許多知名模型像是 BERT 與 GPT-3 的主要架構。

SA 的概念其實非常簡單，主要是透過自身的序列輸入來決定各個時間點的輸入有多大的重要性，經過加權平均之後即可得到輸出。示意圖請見圖 4.3 至圖 4.6，<sup>4</sup> 透過  $W^q$  和  $W^k$  兩個矩陣，可以把各個時間點的輸入轉化成  $q^i$  和  $k^j$ ，透過內積

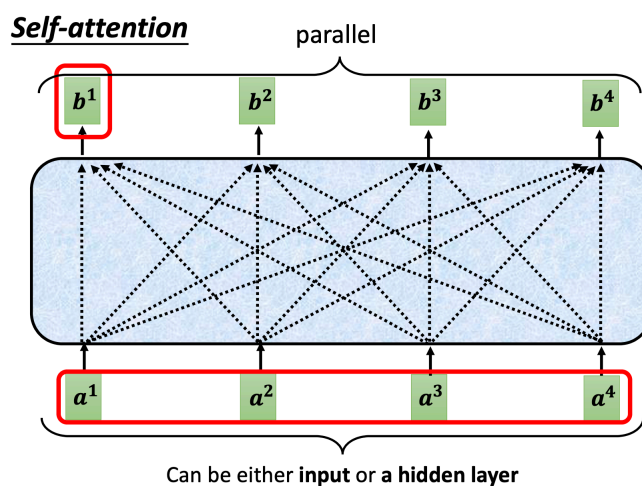


Figure 3.3: Self-Attention 示意圖-1

$q^i, k^j$ ，再經過 Softmax 函數的轉換，即可得到權重  $\alpha'_{i,j}$ ，最後透過  $W^v$  把各個時間點的輸入轉化成  $v^j$  (value)，將  $v^j$  以  $\alpha'_{i,j}$  為權重加權平均後即可得到輸出  $b^i$  若以數

<sup>4</sup>圖片取自國立臺灣大學電機工程系李弘毅教授 2021 年授課投影片。



**Self-attention**

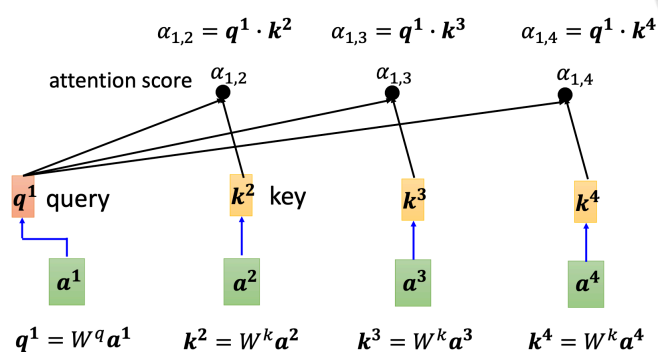


Figure 3.4: Self-Attention 示意圖-2

**Self-attention**

$$\alpha'_{1,i} = \exp(\alpha_{1,i}) / \sum_j \exp(\alpha_{1,j})$$

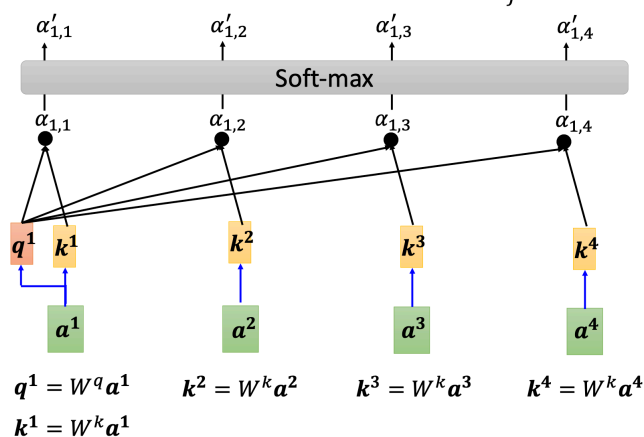


Figure 3.5: Self-Attention 示意圖-3

**Self-attention**

Extract information based on attention scores  $b^1 = \sum_i \alpha'_{1,i} v^i$

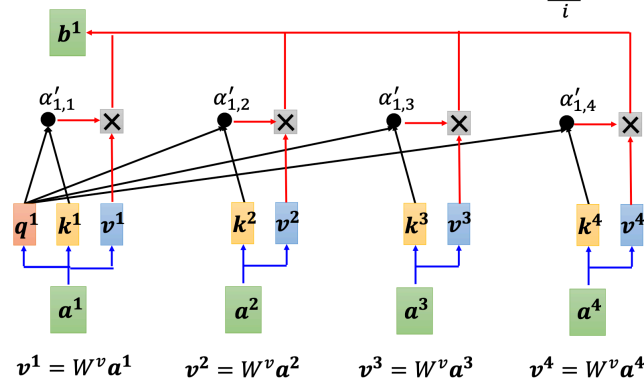


Figure 3.6: Self-Attention 示意圖-4

學式表達則為：<sup>5</sup>

$$Q = W^q a$$

$$K = W^k a$$

$$V = W^v a$$

$$A' = \text{Softmax}(K^T Q)$$

$$b = V A'$$

其中  $a = (a^1, a^2, \dots)$ ,  $b = (b^1, b^2, \dots)$  為 SA 的輸入與輸出， $W_q, W_k, W_v$  為被訓練的參數，維度分別為  $d_q * d_a, d_k * d_a, d_v * d_a$ 。

由於各個時間點的輸入可以透過矩陣乘法同時進行運算，在使用 GPU 的情況下，與循環神經網路相比將可大幅減少運算時間，也由於透過矩陣乘法進行運算，各個時間點的輸入權重將由資料決定，不會因為隔得太過久遠而產生訓練困難，提升了訓練效果，因此成為深度學習模型處理序列資料的主流模型。



---

<sup>5</sup>Softmax() 表示對每一欄 (column) 以 softmax 函數， $a'_{i,j} = \frac{e^{a_{i,j}}}{\sum_j e^{a_{i,j}}}$ ，進行轉換。



## 第四章 實證方法與結果

### 4.1 資料說明及資料集分割

本文所使用的資料來自彭博資料庫 (Bloomberg)，使用美元兌加元及澳幣、英鎊兌美元匯率，<sup>1</sup>所採用之樣本期間為 1981/01/01 到 2020/12/31，頻率為日資料，透過將其分別取對數並進行一階差分，得到匯率變動率作為模型輸入，扣除非交易日共有 10435 筆資料。圖 4.1 至圖 4.3 為原始資料，而圖 4.4 至圖 4.6 則為處理後作為模型輸入之資料，並分別以藍、橘和綠色標示訓練、驗證和測試集樣本。

Table 4.1: 變數說明

變數	變數定義
USDCAD	美元兌加元匯率
AUDUSD	澳幣兌美元匯率
GBPUSD	英鎊兌美元匯率
diff_log_USDCAD	美元兌加元匯率取對數後一階差分
diff_log_AUDUSD	澳幣兌美元匯率取對數後一階差分
diff_log_GBPUSD	英鎊兌美元匯率取對數後一階差分

<sup>1</sup>A 貨幣兌 B 貨幣表示一單位 A 貨幣可換多少單位之 B 貨幣

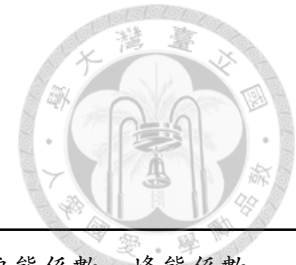


Table 4.2: 變數敘述統計

	平均值	標準差	最小值	最大值	偏態係數	峰態係數	ADF <sup>1</sup>
USDCAD	0.8011	0.0993	0.6198	1.0865	0.5507	-0.3562	-1.87
AUDUSD	0.7785	0.1387	0.479	1.189	0.6367	0.2973	-3.05**
GBPUSD	1.5973	0.2027	1.052	2.427	0.3691	0.4231	-3.92***
diff_log_USDCAD	-0.000006	0.004435	-0.032518	0.039808	-0.1047	5.0084	-104.74***
diff_log_AUDUSD	-0.000041	0.007080	-0.105536	0.082766	-0.7152	13.8020	-104.16***
diff_log_GBPUSD	-0.000053	0.006232	-0.083955	0.045885	-0.3288	6.1417	-98.75***

<sup>1</sup> ADF 表示 Augmented Dickey-Fuller 具截距項單根檢定之檢定統計量，其中增廣項的最適落後期數由 BIC 決定。\*、\*\* 和 \*\*\* 分別表示在 10%、5% 和 1% 的顯著水準下拒絕虛無假設。

訓練－驗證－測試集的分割方式 (Train - Valid - Test split) 為 7:2:1，訓練集資料取 1981/01/01 號至 2008/12/31 號約 7000 筆資料 (約 70% 總樣本數) 對模型參數進行訓練，驗證集資料取 2009/01/01 至 2016/12/31 約 2000 筆資料 (約 20% 總樣本數) 調整超參數 (Hyper-parameter)，測試集資料取 2017/01/01 至 2020/12/31 約 1000 筆資料 (約 10% 總樣本數) 進行模型間的比較。為了減少比較時的不一致性，除了將月資料改成日資料以增加樣本數外，前述採用的匯率及資料集分割方式皆與 Yilmaz and Arabaci (2021) 相同。

## 4.2 研究方法

本節將詳細敘述預測模型的架構、超參數的決定、最佳化器 (Optimizer)、損失函數 (Loss Function) 及模型比較準則的選擇。首先，依照實證模型章節所述，模型的預測將分為線性與非線性兩個部分，分別使用 ARIMA 模型和 LSTM 及 SA，先以 ARIMA 去配適資料，再以 LSTM 及 SA 捕捉剩餘殘差。

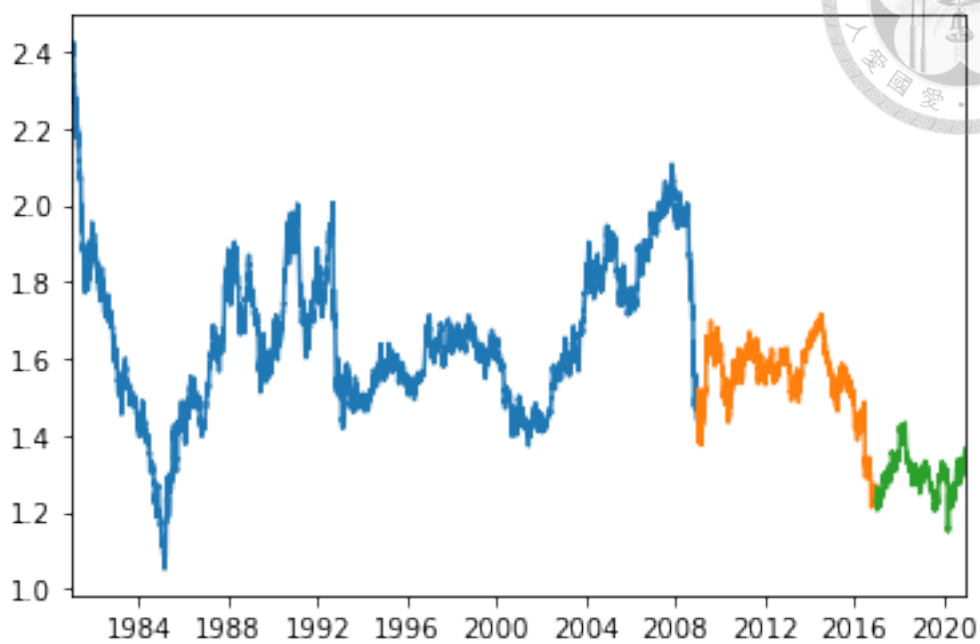


Figure 4.1: 美元兌加元匯率

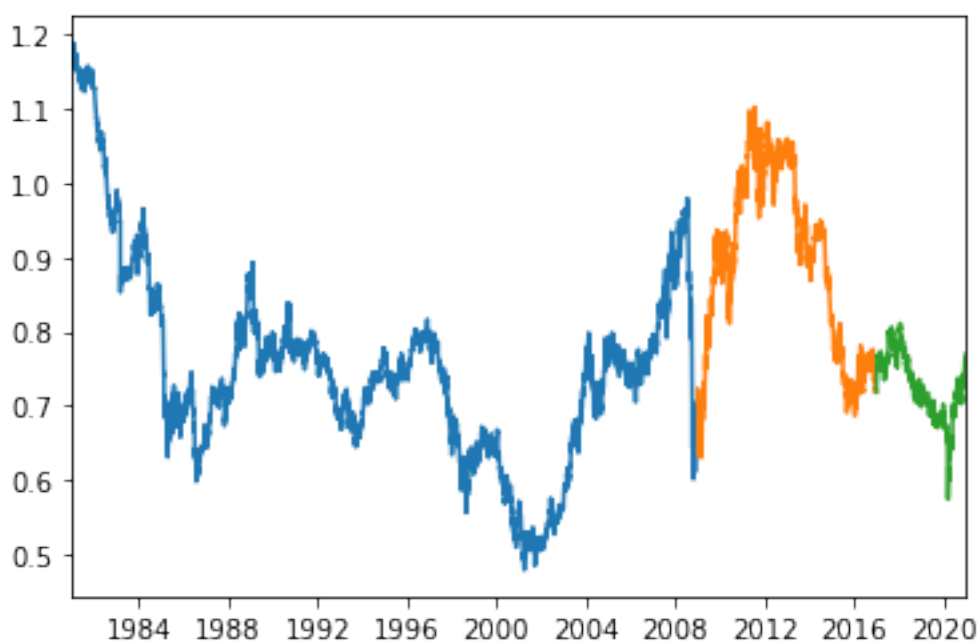


Figure 4.2: 澳幣兌美元匯率

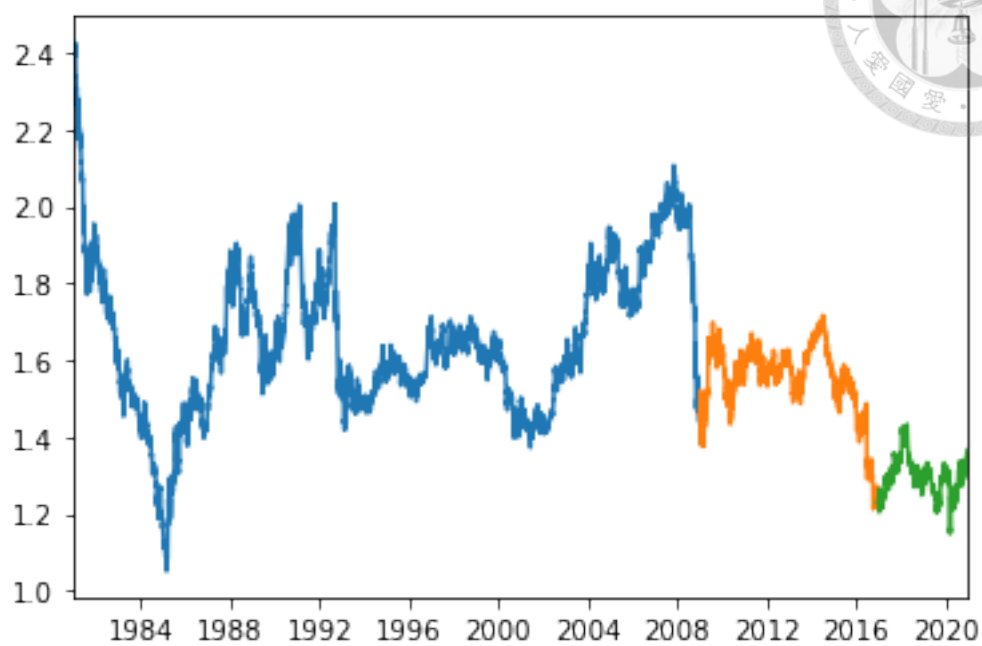


Figure 4.3: 英鎊兌美元匯率

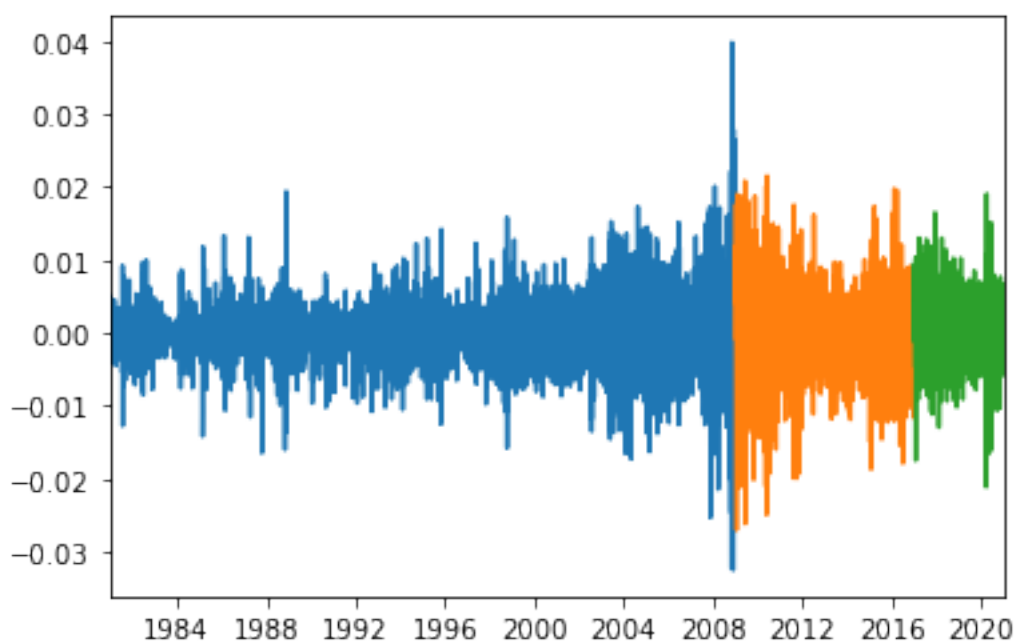


Figure 4.4: 美元兌加元匯率取對數後一階差分

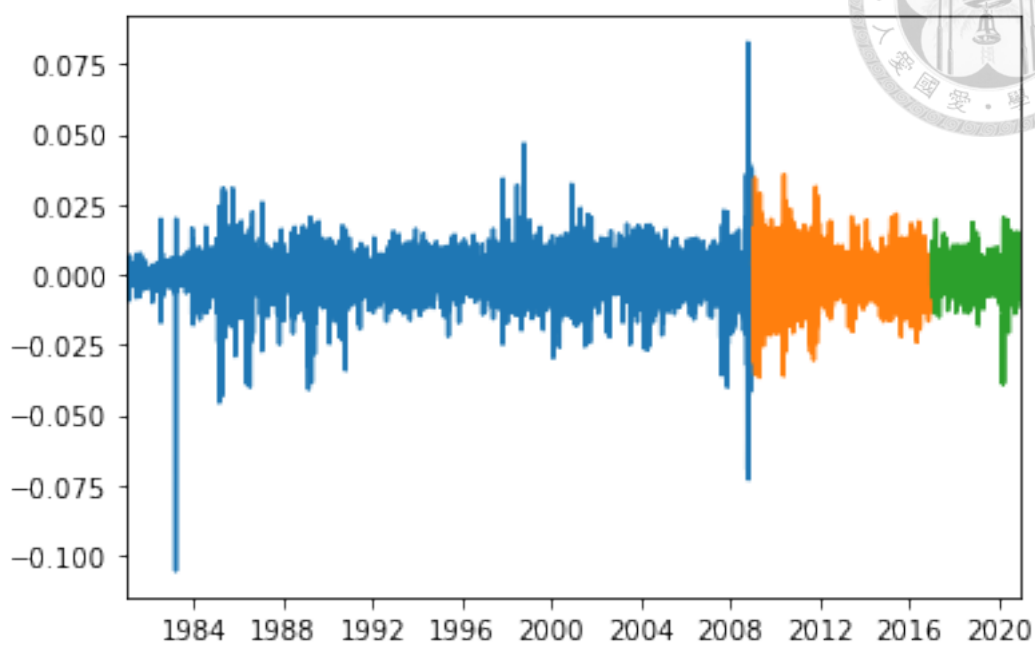


Figure 4.5: 澳幣兌美元匯率取對數後一階差分

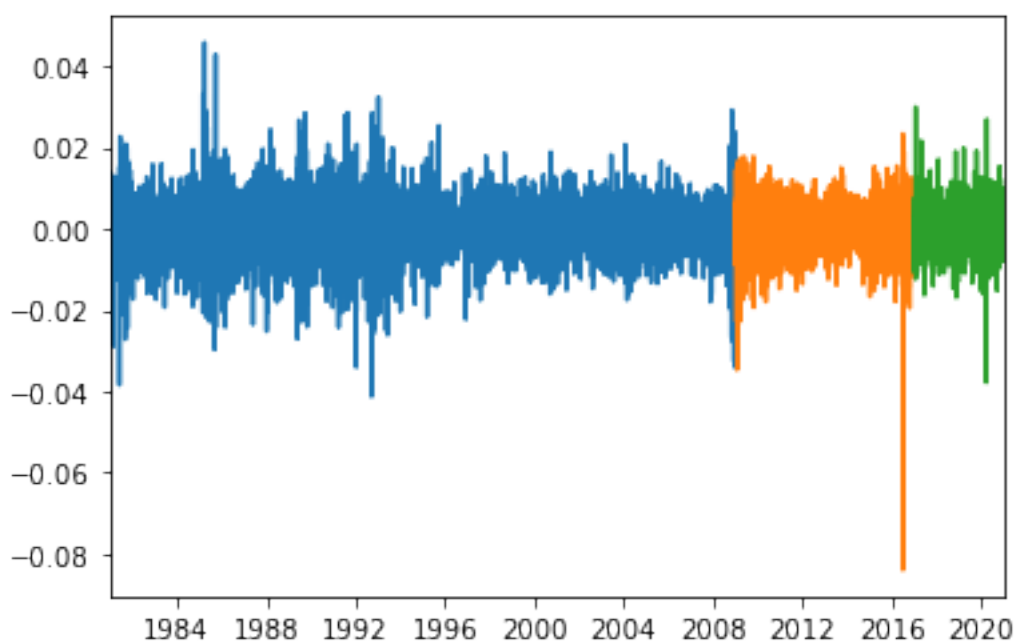


Figure 4.6: 英鎊兌美元匯率取對數後一階差分

ARIMA 模型的部分，我們會分別以滾動視窗 (Rolling Window) 與遞迴視窗 (Recursive Window) 決定拿來配適的資料範圍 (圖 4.7)，再以貝氏資訊準則 (Bayesian Information Criterion, BIC) 來決定最適的參數  $(p, d, q)$ ，並以此預測模型的線性部分及決定殘差。具體來說，滾動視窗和遞迴視窗在形成第一筆預測值時，都會使用全部的訓練集資料來決定參數和預測值，但在形成第二筆預測值時，滾動視窗會排除第一筆訓練資料並加入第一筆驗證集資料，而遞迴視窗則是把第一筆驗證集資料與全部訓練集資料合併；形成第三筆預測值時，滾動視窗會排除第一、二筆訓練資料並加入第一、二筆驗證集資料，而遞迴視窗則是把第一、二筆驗證集資料與全部訓練集資料合併，第四筆及後續預測值就依此類推。



Figure 4.7: 滾動與遞迴視窗示意圖

而 LSTM 及 SA 的部分，當前述 ARIMA 預測完之後，我們將殘差照上一小節的方式分成訓練—驗證—測試集，以此為標籤 (Label) 來訓練非線性部分的深度學習模型。具體來說，我們會使用訓練集資料來訓練模型參數，再以驗證集資料



來選取超參數，<sup>2</sup>其中超參數包括輸入序列長度 (Length of Input Sequence)、迭代次數 (Number of Epochs)、隱藏層數 (Number of Hidden Layers)、隱藏層神經元數 (Number of Cells or Hidden Sizes) 和批量 (Batch Size)；選取超參數的方式為網格搜索 (Grid Search)，超參數搜尋範圍於表 4.3 呈現：<sup>3</sup>

Table 4.3: 超參數網格搜索範圍

輸入序列長度 (Length of Input Sequence) :	5, 10, 20, 60.
迭代次數 (Number of Epochs) :	5, 10, 15, 20, 25, 30.
隱藏層數 (Number of Hidden Layers) :	1, 2, 3, 4.
隱藏層神經元數 (Number of Cells or Hidden Sizes) :	25, 50, 100, 200, 400, 800, 1600.
批量 (Batch Size) :	16, 32, 64, 128, 256.

訓練時使用的最佳化器和損失函數則採用文獻上針對迴歸問題常使用的 Adam 最佳化器和均方誤差 (Mean Squared Error, MSE)。MSE 計算公式如下：<sup>4</sup>

$$MSE = \frac{1}{N} \sum_{t=1}^N (e_t - \hat{e}_t)^2$$

模型比較準則方面，參考 Yilmaz and Arabaci (2021)，我們決定使用均方根誤差 (Root Mean Squared Error, RMSE)、平均絕對誤差 (Mean Absolute Error, MAE) 及方向準確度 (Direction Accuracy, DA) 作為模型比較時使用的準則。計算公式分別如下：<sup>5</sup>

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t|$$

<sup>2</sup>照 Vaswani et al. (2017) 原始論文設定， $d_q = d_k = d_v$ ；為了方便對照 LSTM 與 SA， $d_v$  將等於隱藏層神經元數。

<sup>3</sup>主要參考 Dautel et al. (2020)。

<sup>4</sup>其中  $N$  為批量或驗證集樣本數； $e_t$  和  $\hat{e}_t$  分別為 ARIMA 模型實際預測誤差及深度學習預測值。

<sup>5</sup>其中  $N$  為測試集樣本數， $y_t$  和  $\hat{y}_t$  分別為實際匯率變動率及混合模型預測值； $I[x]$  為指示函數 (Indicator function)，即  $I[x] = 1$  如果  $x > 0$ ，反之  $I[x] = 0$ 。



$$DA = \frac{1}{N} \sum_{t=1}^N I[y_t \cdot \hat{y}_t > 0] \cdot 100$$

### 4.3 實證結果

根據前一節所述之資料處理、模型訓練和超參數調整方法，及模型比較準則的說明，我們將結果整理於表 4.4。

Table 4.4: 各模型匯率預測結果

	ARIMA-LSTM	ARIMA-SA	RW
遮迴視窗			
加元			
MAE	0.003217	0.003229	<b>0.003217</b>
RMSE	0.004367	0.004372	<b>0.004365</b>
DA	49.62	48.28	<b>50.0</b>
澳幣			
MAE	0.004206	0.004192	<b>0.004163</b>
RMSE	0.005716	0.005677	<b>0.005656</b>
DA	<b>50.67</b>	47.80	50.0
英鎊			
MAE	0.004224	0.004226	<b>0.004203</b>
RMSE	0.005707	0.005711	<b>0.005688</b>
DA	48.28	48.28	<b>50.0</b>
滾動視窗			
加元			
MAE	0.003217	0.003229	<b>0.003217</b>
RMSE	0.004367	0.004372	<b>0.004365</b>
DA	49.43	48.28	<b>50.0</b>
澳幣			
MAE	0.004264	0.004190	<b>0.004163</b>
RMSE	0.005780	0.005676	<b>0.005656</b>
DA	<b>50.67</b>	47.80	50.0
英鎊			
MAE	0.004222	0.004223	<b>0.004203</b>
RMSE	0.005705	0.005709	<b>0.005688</b>
DA	48.28	48.56	<b>50.0</b>

粗體字表示最佳模型比較結果。

我們可以看到，改用自注意力機制的 ARIMA-SA 模型的整體預測能力較 ARIMA-LSTM 模型來得差，除了在澳幣預測時均方根誤差及平均絕對誤差較低，和滾動視窗下預測英鎊的方向準確度較高外，其他貨幣別和比較準則都得到較差的結果，顯示自注意力機制在日匯率預測方面的表現不像在自然語言處理相關的時序任務上優異，平行化運算帶來的運算速度提升伴隨著預測表現的下降。甚至，不同於 Yilmaz and Arabaci (2021) 所顯示的結果，在將月資料換成日資料後，除了澳幣在方向準確率的衡量上略優於隨機漫步模型，ARIMA-LSTM 模型



的整體預測能力是劣於隨機漫步模型的，這表示在日匯率預測方面，不論是使用長短期記憶或是自注意力機制，可能都還是無法打敗隨機漫步模型。

Table 4.5: 各模型殘差分析

	加元	澳幣	英鎊
Random Walk	5.62	34.75***	29.84***
遮迴視窗			
ARIMA	53.55***	15.11	8.33
ARIMA-LSTM	6.04	38.62***	28.50***
ARIMA-SA	7.12	34.69***	29.10***
滾動視窗			
ARIMA	53.56***	15.11	8.36
ARIMA-LSTM	6.05	36.58***	28.61***
ARIMA-SA	7.12	34.68***	29.22***

表格內為各模型殘差使用 Ljung-Box 檢定之檢定統計量，落後期數為 12 期。  
\*、\*\* 和 \*\*\* 分別表示在 10%、5% 和 1% 的顯著水準下拒絕虛無假設。

而若是對隨機漫步模型、線性模型 (ARIMA) 與混合模型 (ARIMA-LSTM 和 ARIMA-SA) 的殘差進行檢定 (Ljung-Box 檢定)，可以發現澳幣和英鎊使用隨機漫步模型時就達到殘差獨立的情況，進一步使用線性模型時反而使殘差不獨立，加上了非線型模型的部分才有所改善。反之，加元使用隨機漫步模型時殘差並未獨立，進一步使用線性模型大幅改善了殘差獨立性，加上了原本預期要捕捉非線性特性的類神經網路後，卻得到反效果。



## 第五章 結論

本文的研究為匯率預測，目標係比較以長短期記憶及自注意力機制建立的模型與隨機漫步模型在匯率預測方面的表現差異，主要參考 Yilmaz and Arabaci (2021)，將匯率的預測拆分成線性和非線性兩個部分，分別使用整合移動平均自迴歸模型和自注意力機制分別對美元兌加元及澳幣、英鎊兌美元匯率之報酬率進行預測，並將此模型與 Yilmaz and Arabaci (2021) 研究中的 ARIMA-LSTM 模型及隨機漫步模型進行比較。

### 5.1 研究結論

在決定了模型後，我們先針對資料依模型需求進行前置處理並劃分訓練—驗證—測試集，在訓練及驗證集上嘗試不同的架構並記錄最好的模型參數與結構，並以此訓練模型，最後衡量於測試集。本研究顯示，改用自注意力機制的 ARIMA-SA 模型的預測能力較 ARIMA-LSTM 模型來得差，自注意力機制在匯率預測方面無法得到好於長短期記憶機制的結果。雖然如此，從殘差分析來看，若是原先模型的殘差不獨立，類神經網路是有能力捕捉殘差中的非線性特性的。此外，不同於 Yilmaz and Arabaci (2021) 所顯示的結果，ARIMA-LSTM 模型的預測能力亦劣於隨機漫步模型。考慮到本研究與 Yilmaz and Arabaci (2021) 所使用資料頻率的差異，這可能表示長期策略較短期策略更適用於匯率交易。

## 5.2 研究建議



最後，本研究還有某些可以改進的地方。第一，受限於運算資源及其速度，本研究的超參數搜尋範圍可能恰巧沒有涵蓋到最佳的超參數。第二，隨著深度學習技術不斷地演進，未來可能會出現更多值得研究的類神經網路模型架構。第三，爲了方便與 Yilmaz and Arabaci (2021) 比較，我們沒有使用除了歷史匯率報酬率以外的資料，如公債利率、投資人情緒指標、財經新聞的文本分析等等，若納入這些資料，也有可能得到較佳的結果。綜上所述，搜尋更大範圍的超參數、嘗試更新的模型架構，以及納入更多變數，將是未來研究可能的努力方向。



## 參考文獻

- Alexander, Don and Lee R Thomas III (1987), “Monetary/Asset models of exchange rate determination: How well have they performed in the 1980’s?”, *International Journal of Forecasting*, 3(1), 53–64.
- Alvarez-Diaz, Marcos (2008), “Exchange rates forecasting: local or global methods?”, *Applied Economics*, 40(15), 1969–1984.
- Dautel, Alexander Jakob, Wolfgang Karl Härdle, Stefan Lessmann, and Hsin-Vonn Seow (2020), “Forex exchange rate forecasting using deep recurrent neural networks”, *Digital Finance*, 2(1), 69–96.
- Dodevski, Aleksandar, Natasa Koceska, and Saso Koceski (2018), “Forecasting exchange rate between macedonian denar and euro using deep learning”, *Journal of Applied Economics and Business*, 6(2), 50–61.
- Engle, Charles and James D Hamilton (1990), “Long swings in the dollar: Are they in the data and do markets know it”, *American Economic Review*, 80(4), 689–713.
- Giles, C Lee, Steve Lawrence, and Ah Chung Tsoi (2001), “Noisy time series prediction using recurrent neural networks and grammatical inference”, *Machine Learning*, 44(1), 161–183.

Hochreiter, Sepp and Jürgen Schmidhuber (1997), “Long short-term memory”, *Neural Computation*, 9(8), 1735–1780.



Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989), “Multilayer feedforward networks are universal approximators”, *Neural Networks*, 2(5), 359–366.

Kiani, Khurshid M and Terry L Kastens (2008), “Testing forecast accuracy of foreign exchange rates: Predictions from feed forward and various recurrent neural network architectures”, *Computational Economics*, 32(4), 383–406.

Lai, Kin Keung, Lean Yu, Shouyang Wang, and Wei Huang (2006), “Hybridizing exponential smoothing and neural network for financial time series predication”, in *Computational Science – ICCS 2006*, edited by Vassil N. Alexandrov, Geert Dick van Albada, Peter M. A. Sloot, and Jack Dongarra, Berlin, Heidelberg: Springer Berlin Heidelberg.

Meese, Richard A. and Kenneth Rogoff (1983), “Empirical exchange rate models of the seventies: Do they fit out of sample?”, *Journal of International Economics*, 14(1), 3–24.

Qu, Yaxin and Xue Zhao (2019), “Application of LSTM neural network in forecasting foreign exchange price”, *Journal of Physics: Conference Series*, 1237(4), 042036.

Ranjit, Swagat, Shruti Shrestha, Sital Subedi, and Subarna Shakya (2018), “Comparison of algorithms in foreign exchange rate prediction”, in *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*.

Rogoff, Kenneth S and Vania Stavrakeva (2008), “The continuing puzzle of short horizon exchange rate forecasting”, Working Paper 14071, National Bureau of Economic Research.

Rossi, Barbara (2013), “Exchange rate predictability”, *Journal of Economic Literature*, 51(4), 1063–1119.



Sezer, Omer Berat, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu (2020), “Financial time series forecasting with deep learning : A systematic literature review: 2005—2019”, *Applied Soft Computing*, 90, 106181.

Tenti, Paolo (1996), “Forecasting foreign exchange rates using recurrent neural networks”, *Applied Artificial Intelligence*, 10(6), 567–582.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin (2017), “Attention is all you need”, in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, vol. 30, Curran Associates, Inc.

Wolff, Christian CP (1988), “Models of exchange rates: A comparison of forecasting results”, *International Journal of Forecasting*, 4(4), 605–607.

Wu, Yungao and Jianwei Gao (2018), “AdaBoost-based long short-term memory ensemble learning approach for financial time series forecasting”, *Current Science*, 115(1), 159–165.

Yilmaz, Firat Melih and Ozer Arabaci (2021), “Should deep learning models be in high demand, or should they simply be a very hot topic? A comprehensive study for exchange rate forecasting”, *Computational Economics*, 57(1), 217–245.