

# 臺北捷運旅運量經濟分析及預測

張哲瑋

壹、	前言 .....	2
一、	研究動機與目的 .....	2
貳、	研究方法 .....	2
一、	HP – Filter .....	2
二、	STL 分解 ( Seasonal and Trend decomposition using Loess ) .....	2
三、	ARIMA ( AutoRegressive Integrated Moving Average ) .....	3
四、	ETS Model ( Error-Trend-Seasonality ) .....	3
參、	資料集 .....	4
一、	資料來源 .....	4
二、	數據型態 .....	4
三、	誤差指標 ( Error Metrics ) .....	5
肆、	實證分析 .....	6
一、	HP – Filter .....	6
二、	STL Decomposition .....	6
三、	ARIMAM .....	7
四、	ETS Model .....	8
伍、	結論 .....	9
陸、	參考文獻 .....	9
柒、	參考資料 .....	10

## 壹、前言

### 一、研究動機與目的

在臺灣，大眾交通運輸已融入我們日常生活的不可或缺之一，尤其在大臺北地區，臺北捷運系統已深刻改變現代居民的生活模式。這使得對近 20 年臺北捷運的運旅量進行深入研究顯得至關重要，以揭示是否存在明顯的趨勢和季節性變化，同時透過這些數據推算未來臺北捷運的運旅量。

## 貳、研究方法

### 一、HP - Filter

HP-Filter 是由 Hodrick 和 Prescott 於 1997 年提出的一種時間序列分析方法，主要用於區分時間序列中的長期趨勢和循環波動。該方法假設任何時間序列  $y_t$  可以被分解為長期趨勢  $\tau_t$  和循環波動  $c_t$  兩部分。HP 濾波器的目標是通過以下最優化模型來找到  $\tau_t$  的最優解：

$$\min_{\tau_t} \sum_{t=1}^T (y_t - \tau_t)^2 + \lambda(\tau_{t+1} - 2\tau_t + \tau_{t-1})^2 \quad 1$$

其中， $T$  是時間序列的長度； $\lambda$  是平滑參數，第一項為剷除長期趨勢  $\tau_t$ ，得到循環波動  $c_t$ ，第二項是對長期趨勢進行平滑處理，降低長期趨勢的變動。通過調整平滑參數  $\lambda$ ，可以控制對趨勢的平滑程度，通常建議年資料參數值  $\lambda = 100^2$ ；季資料參數值  $\lambda = 1600^3$ ；月資料參數值  $\lambda = 14400^4$ 。

### 二、STL 分解 (Seasonal and Trend decomposition using Loess) <sup>5</sup>

STL 利用局部加權回歸 (Loess) 進行平滑處理，以分解時間序列數據成為季節性 (seasonality,  $S_v$ )、趨勢 (trend,  $T_v$ ) 和殘差 (Residuals,  $R_v$ )。

loess (Locally estimated scatterplot smoothing)，它通過對每一點  $x$  附近的點進行加權，進行多項式擬合來估計函數  $\hat{g}(x)$  的值。在這個過程中，每一點的權重是基於其距離  $x$  的遠近而定，距離越近的點權重越大，這樣可以保證對  $x$  點的預測更加依賴於其周圍的數據點。

---

<sup>1</sup> 此方程式根據 Alfaro, R., & Drehmann, M. (2023). The Holt-Winters filter and the one-sided HP filter: A close correspondence. *Economics Letters*, 222, 110925.

<sup>2</sup> 此參數設定根據 Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1-16.

<sup>3</sup> 此參數設定根據 Backus, D. K., & Kehoe, P. J. (1992). International evidence on the historical properties of business cycles. *The American Economic Review*, 864-888.

<sup>4</sup> 此參數設定根據 Zarnowitz, V., & Ozyildirim, A. (2006). Time series decomposition and measurement of business cycles, trends and growth cycles. *Journal of Monetary Economics*, 53(7), 1717-1739.

<sup>5</sup> STL 分解參考 Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *J. Off. Stat.*, 6(1), 3-73.

再來，我們會進行一個迭代迴圈，內圈尋找最佳的趨勢（trend component,  $T_v$ ）與季節性（seasonal component,  $S_v$ ），而外圈有助於分解穩定性，在執行一個外圈時，會先從內圈得到 $T_v$ 與 $S_v$ 的估計值，再計算出殘差（Residuals,  $R_v$ ）：

$$R_v = Y_v - T_v - S_v$$

這些殘差（ $R_v$ ）是評估數據是否偏離模型預測的一種指標。如果殘差大，那可能意味著該數據是一個異常值或者模型未能捕捉到的複雜性。STL 使用這些殘差來調整權重，具體方法是在下一次外圈迭代進行時，將會減少那些具有大殘差的數據之權重。這樣做可以降低這些點在趨勢和季節性成分估計中的影響。

這個權重調整過程在 STL 算法的每個外圈重複進行，最終分解結果的穩定性。通過迭代這個過程，使得我們最終分解的結果考慮數據整體趨勢和季節性，又能適當的處理異常值。

### 三、ARIMA（AutoRegressive Integrated Moving Average）

ARIMA 是一種時間序列分析的方法，用於預測未來數值。ARIMA 模型結合了自回歸（AR）和移動平均（MA）兩種元素，並通過整合（I，即差分過程）來處理非定態時間序列<sup>6</sup>數據。ARIMA 模型的一般表示法是 ARIMA(p, d, q)，其中：

- p：自回歸（AR）的階數，表示模型中使用多少個先前時間點的數值進行預測。
- d：整合（I）的次數，表示需要對原始數據進行多少次差分，使其變得趨於穩定。
- q：移動平均（MA）的階數，表示模型中使用多少個先前時間點的殘差進行預測。

### 四、ETS Model（Error-Trend-Seasonality）

ETS 模型是一種時間序列分析的工具，用以將時間序列數據進行拆解以進行更深入的分析。通常會包含誤差（Error）、趨勢（Trend）及季節性（Seasonality），而這三個參數可以透過不同的方式組成，形成不同種類的 ETS 模型。

對於 ETS 模型，考慮到不同的參數設定是很重要的，而這些參數之間的組合方式可以是加法（add）或是乘法（mul）。這種不同的組合方式將直接影響模型的結構和性能。具體而言，ETS 模型的數學表示分為加法和乘法兩種情況：

ETS (A, A<sub>d</sub>, A)：

$$\hat{y}_t = \ell_{t-1} + \phi \cdot b_{t-1} + s_{t-m}$$

$$\ell_t = \ell_{t-1} + \phi \cdot b_{t-1} + \alpha \cdot \varepsilon_t$$

$$b_t = \phi \cdot b_{t-1} + \beta \cdot \varepsilon_t$$

$$s_t = s_{t-m} + \gamma \cdot \varepsilon_t$$

ETS (M, M<sub>d</sub>, M)：

$$\hat{y}_t = \ell_{t-1} \cdot b_{t-1}^\phi \cdot s_{t-m}$$

$$\ell_t = \ell_{t-1} \cdot b_{t-1}^\phi \cdot (1 + \alpha \cdot \varepsilon_t)$$

$$b_t = b_{t-1}^\phi (1 + \beta \cdot \varepsilon_t)$$

$$s_t = s_{t-m} (1 + \gamma \cdot \varepsilon_t)$$

<sup>6</sup> 非定常時間序列：具有趨勢、季節性或是週期之時間序列。

以下為公式中的變數意涵：

- $\ell_t$ ：在  $t$  時的水準（Level）。
- $b_t$ ：在  $t$  時的趨勢（Trend）。
- $s_t$ ：在  $t$  時地季節性（Seasonality）。
- $\varepsilon_t$ ：在  $t$  時的誤差（Error）。
- $\Phi$ ：趨勢的阻尼係數（Damped Trend），用以控制趨勢衰減的速率。
- $\alpha$ ：水準的平滑參數。
- $\beta$ ：趨勢的平滑參數。
- $\gamma$ ：季節性的平滑參數。
- $t$ ：當期時間點。
- $m$ ：季節性週期長度。

## 參、 資料集

### 一、 資料來源

本研究的資料來源為臺北大眾捷運股份有限公司提供的全系統旅運量統計數據。資料涵蓋的時間範圍自西元 2003 年 1 月 1 日至 2022 年 12 月 31 日，總計 20 年的日資料數據，目的在進行深入的分析並建立預測模型。

在模型的建立過程中，我們使用了 2003 年 1 月 1 日至 2022 年 12 月 31 日的資料來進行模型的訓練和調整。而後，我們將模型應用於 2023 年 1 月 1 日至 12 月 31 日的數據，以評估其預測性能並確保模型的準確性。

### 二、 數據型態

1. 月資料（Table 1）
2. 季資料（Table 2）
3. 年資料（Table 3）
4. 周平均資料（Table 4）
5. 月平均資料（Table 5）

	旅客量
2003-01-01	28014216
2003-02-01	26241269
2003-03-01	29040656
2003-04-01	25791169
2003-05-01	18647030
⋮	⋮

Table 1 月資料

	旅客量
2003-01-01	83296141
2003-04-01	65151747
2003-07-01	79972072
2003-10-01	87769168
2004-01-01	85420247
⋮	⋮

Table 2 季資料

	旅客量
2003-01-01	316189128
2004-01-01	350141956
2005-01-01	360729803
2006-01-01	384003220
2007-01-01	416229685
⋮	⋮

Table 3 年資料

	旅客量
Monday	1558892
Tuesday	1586923
Wednesday	1608753
Thursday	1611748
Friday	1719971
⋮	⋮

Table 4 周平均資料

	旅客量
01-01	46686325
02-01	42142431
03-01	49578836
04-01	45484570
05-01	45235277
⋮	⋮

Table 5 月平均資料

### 三、 誤差指標（Error Metrics）

BIC（Bayesian Information Criterion），是一種模型選擇準則，用於在統計建模中比較不同的統計模型。BIC 的計算方式如下：

$$\text{BIC} = -2 \cdot \ln(\mathcal{L}) + k \cdot \ln(n)$$

- $\mathcal{L}$ ：模型的最大概似函數值（likelihood）。
- $k$ ：模型的參數個數。
- $n$ ：樣本數。

在模型選擇中，我們會比較不同模型的 BIC 值，選擇 BIC 最小的模型作為最佳模型。較小的 BIC 值表示在概似函數值相對較大的同時，考慮到複雜性的懲罰，模型的總體表現更好。

RMSE（Root Mean Squared Error），是用於評估預測模型在連續變數預測中的準確性的一種指標。RMSE 是預測誤差的平方根的平均值，它度量了實際觀測值和模型預測值之間的平均差異。RMSE 的計算方式如下：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

- $n$ ：樣本數。
- $Y_i$ ：實際值。
- $\hat{Y}_i$ ：模型的預測值。

RMSE 的值越小，表示模型的預測與實際觀測值之間的差異越小，模型的準確性越高。然而，RMSE 對於極端值敏感，因為它使用了誤差的平方，這會使得大的誤差在計算中得到放大。

## 肆、 實證分析

### 一、 HP - Filter

HP-Filter 的應用結果顯示在三張圖中，分別為年資料、季資料和月資料的趨勢。這三個數據集都呈現出了自 2002 年至 2019 年的上升趨勢，但在 2020 年迎來了一個明顯的下降。這主要是受到 COVID-19 疫情的影響，該疫情對人們的生活習慣產生了巨大的影響。

通過觀察年資料（Figure 1），我們可以看到整體上僅有輕微的下降趨勢。經過年平均處理後，原始數據中的季節性和短期趨勢已被消除。因此，我們可以將 COVID-19 視為一個短期波動，而長期趨勢仍然呈現上升。

從季資料（Figure 2）和月資料（Figure 3）來看，我們認為循環波動中包含季節性。因此，接下來我們將使用 STL（Seasonal and Trend decomposition using Loess）進行更全面的分解，以更深入地了解數據中的季節性成分。

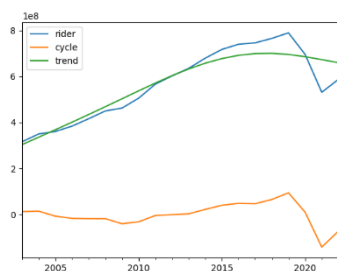


Figure 1 HP-Filter（年資料）

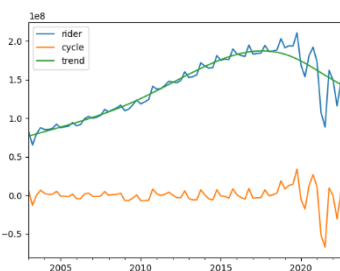


Figure 2 HP-Filter（季資料）

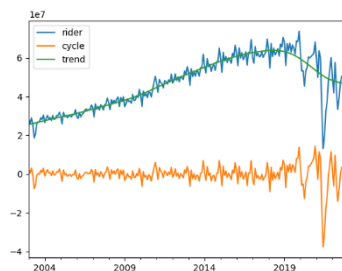


Figure 3 HP-Filter（月資料）

### 二、 STL Decomposition

STL 分解的應用結果顯示在下列兩張圖中，分別為季資料（Figure 4）和月資料（Figure 5）的拆解，可以從圖中的 Season 得知臺北捷運旅運量是有明顯的季節性，同時，我們還能夠辨識出在 2020 年至 2022 年期間，COVID-19 疫情對捷運旅運量產生的極端值，這反映了疫情對大眾交通使用的影響。

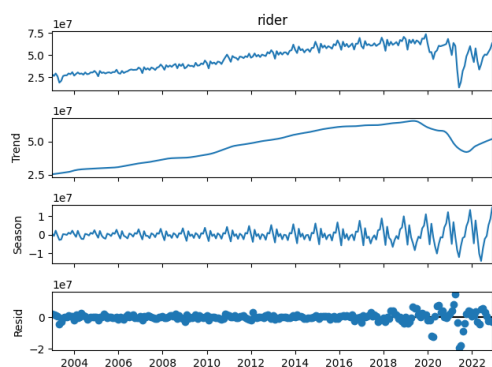


Figure 4 STL 分解（季資料）



Figure 5 STL 分解（月資料）

### 三、ARIMAM

通過 STL 季節性分解和 ARIMA 時間序列建模，我們首先捕捉時間序列的季節性結構，然後專注於預測去季節性的趨勢。這種方法結合了季節性和非季節性特徵，提高了對未來趨勢的預測準確性。

ARIMA 之參數設定，我們使用 `auto.arima()` 尋找最適參數，我們透過此函數可以得知，對於月資料和季資料最適參數皆為 (4, 1, 1)，也就是說我們對於月資料和季資料之估計模型皆採用 ARIMA(4, 1, 1)。以下表表格為預測結果、實際資料以及 RMSE。

日期	預測值	實際值
2023-01-01	54487331	52502386
2023-02-01	44944433	56091010
2023-03-01	57273890	62605487
2023-04-01	48730768	56684568
2023-05-01	39391577	60510291
2023-06-01	35462707	54987194
2023-07-01	42119104	58046189
2023-08-01	47280715	58839358
2023-09-01	48818138	58555890
2023-10-01	54586796	61214212
2023-11-01	57965164	62840839
2023-12-01	63983764	67308154
RMSE：9,706		

Table 6 月資料 ARIMA(4, 1, 1)預測結果



日期	預測值	實際值
2023-01-01	155351758	171198883
2023-04-01	123084817	172182053
2023-07-01	135000723	175441437
2023-10-01	176847773	191363205
RMSE：21,114		

Table 7 季資料 ARIMA(4, 1, 1)預測結果

#### 四、ETS Model

ETS 模型之參數我們將考慮誤差的加法與乘法；趨勢的加法與乘法；季節性的加法與乘法以及趨勢的阻尼係數，一共 36 種模型，選擇其 BIC 最小值之參數，作為模型設定。在月資料和季資料中，由於 ETS (A,N,M) 的 BIC 最小，故選擇此皆模型作為我們的預測模型。下表表格為預測結果、實際資料以及 RMSE。

日期	預測值	實際值
2023-01-01	57315716	52502386
2023-02-01	51045769	56091010
2023-03-01	59688787	62605487
2023-04-01	54708599	56684568
2023-05-01	54212498	60510291
2023-06-01	51445656	54987194
2023-07-01	54683838	58046189
2023-08-01	55226653	58839358
2023-09-01	53185895	58555890
2023-10-01	57537911	61214212
2023-11-01	57952329	62840839
2023-12-01	63326154	67308154
RMSE：4,278,841		

Table 8 月資料 ETS (A, N, M) 預測結果

日期	預測值	實際值
2023-01-01	152070828	171198883
2023-04-01	132390992	172182053
2023-07-01	144531159	175441437
2023-10-01	172608392	191363205
RMSE : 12,634,723		

Table 9 季資料 ETS (A, N, M) 預測結果

## 伍、 結論

我們可以從 HP-Filter 以及 STL 分解的結果得知，臺北捷運系統的乘客流量展示了一個長期的上升趨勢，並且呈現出顯著的季節性波動，此外，在 2020 年至 2022 年期間，由於 COVID-19 疫情的影響，數據中亦觀察到多個異常值。這些異常值可能反映了疫情對於公共交通使用率的影響，包括隨著疫情波動而起伏的乘客數量。

在採用 ARIMA(4, 1, 1) 及 ETS (A, N, M) 模型進行預測時，即使在進行精細的參數調整後，這兩種模型計算出的季資料和月資料的均方根誤差 (RMSE) 依然相對較大。我們推斷這一結果可能源於數據集中存在異常值的問題，尤其是在 2020 年至 2022 年期間，由於 COVID-19 疫情的影響，這段時間內的數據顯示出與以往年份顯著不同的趨勢和波動，尤其是當目的是要對 2023 年的旅運量進行準確預測。因此，這些極端值對於預測的準確性有著顯著的影響。

經過對季資料和月資料進行詳細的均方根誤差 (RMSE) 比較分析後，顯示 ARIMA (4, 1, 1) 模型在這兩種時間序列上都產生了明顯較低的 RMSE 值，相較於 ETS (A, N, M) 模型。基於這一結果，我們可以合理推斷，在我們臺北捷運旅運量的數據集中，ARIMA (4, 1, 1) 模型在預測準確度上優於 ETS (A, N, M) 模型。這表明 ARIMA (4, 1, 1) 模型更適合捕捉和反映該時間序列的潛在動態和結構特。

## 陸、 參考文獻

Alfaro, R., & Drehmann, M. (2023). The Holt-Winters filter and the one-sided HP filter: A correspondence. *Economics Letters*, 222, 110925.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1-16.

Backus, D. K., & Kehoe, P. J. (1992). International evidence on the historical properties of business cycles. *The American Economic Review*, 864-888.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *J. Off. Stat*, 6(1), 3-73.

Zarnowitz, V., & Ozyildirim, A. (2006). Time series decomposition and measurement of business cycles, trends and growth cycles. *Journal of Monetary Economics*, 53(7), 1717-1739.

## 柒、 參考資料

<https://ws.ndc.gov.tw/Download.ashx?u=LzAwMS9hZG1pbmlzdHJhdG9yLzEwL3JlbGZpbGUvNTY3Mi80MzM1LzAwMDg5MDVfMy5wZGY%3D&n=Y2hhcHRlciAxQDE3MTYyNi4yNDY5NjczMThzMTgzMkAucGRm&icon=.pdf>

<https://ws.ndc.gov.tw/Download.ashx?u=LzAwMS9hZG1pbmlzdHJhdG9yLzEwL3JlbGZpbGUvMC8xMDczNC80NzgZGZGRmMi00NzE4LTQ2NDQtdDdhYi04MzI0NGRkN2RhZDIucGRm&n=MuaZgumWk%2BaVuOWIl%2BmVt%2Bacn%2Bi2qOWLouS8sOioiOS5i%2BeglOepti5wZGY%3D&icon=.pdf>

<https://medium.com/r-%E8%AA%9E%E8%A8%80%E8%87%AA%E5%AD%B8%E7%B3%BB%E5%88%97/r%E8%AA%9E%E8%A8%80%E8%87%AA%E5%AD%B8%E6%97%A5%E8%A8%98-8-%E8%99%95%E7%90%86%E5%AD%A3%E7%AF%80%E6%80%A7%E5%95%8F%E9%A1%8C-%E5%82%85%E7%AB%8B%E8%91%89%E8%AE%8A%E6%8F%9B%E8%88%87stl%E5%88%86%E8%A7%A3-80ff3c56e142>

<https://www.statsmodels.org/stable/generated/statsmodels.tsa.forecasting.stl.STLForecast.html>

<https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima.model.ARIMA.html#statsmodels.tsa.arima.model.ARIMA>

<https://www.statsmodels.org/dev/examples/notebooks/generated/ets.html>

[https://www.statsmodels.org/devel/generated/statsmodels.tsa.exponential\\_smoothing.ets.ETSModel.html#rdfed55d5c7db-1](https://www.statsmodels.org/devel/generated/statsmodels.tsa.exponential_smoothing.ets.ETSModel.html#rdfed55d5c7db-1)