

台灣通膨率預測：運用大數據資料分析*

林馨怡、顏佑銘、葉錦徽**

摘要

本文選擇93個與台灣通貨膨脹相關的總體經濟及貨幣金融變數，構建2000年1月至2023年12月的大數據資料庫，並應用時間序列模型、LASSO家族模型、因子模型，非線性機器學習方法，以及組合模型，預測台灣通膨率。實證結果顯示，在向前一期通膨預測方面，線性與非線性機器學習模型，以及組合模型的預測表現相近，均優於時間序列模型。在短期或中長期通膨率預測方面，非線性機器學習模型預測表現最佳，此結果可能來自重要總體經濟變數與通膨之間潛在非線性關係。再者，觀察預測模型選擇對通膨預測之重要變數，我們發現台灣通膨率深受國際因素及進出口貿易影響，主要貿易對手國相關變數富含通膨預測資訊，故在多個預測模型皆被選為重要變數。此外，國內因素中，貨幣總計數、市場利率、批發業與零售業營業額，以及供給面相關變數對通膨預測亦有貢獻。最後，房市及股市相關變數，以及指數類型變數涵蓋市場重要訊息，也富含通膨預測資訊。

關鍵詞：通膨率、機器學習、隨機森林、因子模型

JEL分類代號：C22，E31

* 本文摘錄自央行委託之研究計畫報告，若有任何錯誤，均由作者負責，不代表中央銀行及作者服務單位之立場。感謝徐之強教授、劉祝安研究員、央行經濟研究處吳懿娟處長及同仁，以及金融業務檢查處同仁提供寶貴意見，我們受益良多。感謝央行經濟研究處陳佩玗專員與政大經濟研究所碩士陶芷敏、賴沛承、謝錚奇的協助。

** 林馨怡為國立政治大學經濟學系教授、顏佑銘為國立政治大學國際經營與貿易學系副教授、葉錦徽為國立中央大學財務金融學系教授。

壹、前言

通貨膨脹是總體經濟重要指標之一，若能了解通貨膨脹走勢，將有助於判斷當前經濟環境及預測未來經濟狀況。目前關於通膨預測的文獻，除了以傳統時間序列模型為基礎的計量模型外，主要分為四大類。第一大類以總體理論模型為預測模型，第二大類則是嘗試將意見調查結果納入通膨預測研究，第三大類則先以多個總體變數建構新指標，包括擴散指標(diffusion index)、因子模型(factor model)，以及動態因子模型(dynamic factor model)等，再接續進行通膨預測。近年來，由於許多大型總體金融數據資料庫的建構，加以機器學習方法蓬勃發展，因此，許多學者將機器學習方法運用於通膨預測，此為第四大類文獻。相關文獻可進一步細分為線性機器學習方法，如：Medeiros and Mendes (2016)，與非線性機器學習方法，如：Inoue and Kilian (2008)、Choudhary and Haider (2012)、Kohlscheen (2021)、Medeiros et al. (2021)、Barkan et al. (2023) 以及 Almosova and Andersen (2023) 等。

前述文獻大都以美國為實證範例，近年來陸續有學者使用各式機器學習方法進行其他國家的通膨預測，舉例而言，Baybuza (2018) 關注俄羅斯之通膨預測，Esquivel Monge (2007) 和 Rodriguez-Vargas (2020) 預測哥斯大黎加通膨，以及 Medeiros et al.

(2016)和 Araujo and Gaglianone (2023)進行巴西的通膨預測。此外，亦有多篇研究關注台灣通膨率預測，包括：陳佩玗(2013)、朱浩榜 (2019)、蕭宇翔與繆維正(2021)、林依伶與陳佩玗(2021)與朱浩榜(2023)等。然而，目前預測台灣通膨率的文獻，較少應用大量總體金融變數建構之大數據資料預測。此外，一旦建立大數據資料後，傳統計量模型不再適用，需應用機器學習方法才能進行預測及分析。本文目的為應用機器學習方法於台灣通膨率預測，期待能提高台灣通膨率預測之準確度，俾有助於央行擬訂適合的貨幣政策。

本文應用總體經濟及金融市場大數據資料，提出多種計量與機器學習方法預測台灣通膨率，並比較預測模型之預測表現。首先，本文選擇93個與台灣通貨膨脹息息相關的總體經濟及貨幣金融變數，建構2000年1月至2023年12月間的大數據資料庫，其中變數分為 11 個類別，包括：景氣指標、就業、民生消費、貨幣金融、股票市場、債券市場、房地產市場、大宗物資、國際收支與進出口、海外因素及情緒指標。接著，我們以時間序列模型為基準模型，並考慮多種機器學習模型進行通膨預測，其中，模型包括：LASSO (least absolute shrinkage and selection operator)家族模型(含脊迴歸

(ridge regression)、LASSO、EN、Adaptive LASSO、Adaptive EN等方法)、因子模型(含主成分分析法及目標因子方法(target factors))、非線性機器學習方法(含套袋樹(bagging tree)與隨機森林(random forest)方法),以及組合預測等模型。預測期間涵蓋向前一期、短期(一年以內),以及中長期(二年),以檢視在不同預測期間下,各預測模型的表現差異。此外,本文利用預測模型挑選變數,識別隱含重要通膨預測資訊的預測變數,使得通膨預測不僅及於不同預測模型間演算方法的比較,更可藉由檢視對預測台灣通膨重要影響的關鍵變數,進一步瞭解影響通膨走勢的傳遞管道。

由向前一期通膨預測之實證結果可得知,LASSO模型家族、非線性機器學習模型及組合模型的預測表現較基準模型佳,而因子模型的預測表現較基準模型差。其中,非線性機器學習模型及組合模型與LASSO模型家族的預測表現差異不大。在非線性機器學習模型中,以套袋樹法表現較佳,組合模型中則以LASSO-套袋樹法表現較佳。此外,LASSO模型家族、非線性機器學習模型,以及組合模型的向前一期通膨預測的樣本外預測走勢與真實通膨率亦步亦趨,皆能準確地預測台灣通膨率。再者,相較於向前一期之預測表現,一年內之短期通膨率預測之預測誤差皆有上升趨勢,其中,非線性機器學習方法預測誤差較小,且明顯改善基準模型預

測表現。關於中長期通膨預測方面,亦以非線性機器學習方法預測表現最佳。綜觀不同預測期間的樣本外多期預測結果可以看出,LASSO模型家族、非線性機器學習模型及組合模型均在多期預測表現佳,其中隨機森林方法在所有多期預測的預測期間下,皆維持穩定且良好的預測表現。

為了進一步瞭解關於大數據資料庫眾多變數預測通膨的能力,本文觀察預測模型在不同預測期間所選擇對通膨預測有解釋能力的重要變數,我們發現,台灣通膨深受國際因素及進出口貿易影響,故主要貿易對手國相關變數大多經選為重要變數,惟就非線性機器學習模型及組合模型而言,此一影響隨預測期間拉長而有所下降。其次,國內因素對通膨預測較有貢獻者,除前一年經季節調整(季調)的CPI年增率外,尚有貨幣總計數、市場利率、批發業與零售業營業額,以及供給面相關變數等。再者,其他國內因素,如房市及股市相關變數、景氣相關指標變數,以及指數類型變數亦經常被選為重要變數,顯示其涵蓋所屬市場的重要訊息,並且富含通膨預測資訊內涵。

綜上所述,在比較不同模型的預測表現後,本文發現應用大數據資料預測台灣通膨率時,無論是短、中、長期預測期間,非線性機器學習方法模型之預測表現皆優於傳統時間序列方法。藉由檢視對預測台灣通膨有重要影響的關鍵變數,我們進一步瞭解對通

膨走勢具有影響力的傳遞管道以及其政策意涵。近期許多研究指出，重要總體經濟與金融變數之間存在非線性關係，由於非線性機器學習方法擅長捕捉經濟體系中潛藏的非線

性特徵，從而廣泛應用於通膨預測，與本文的實證結果一致，因此，此一研究方法可供央行未來進行通膨預測時參考，以利後續政策執行，並達到央行經營目標。

貳、研究方法

本文使用之通膨率預測模型如下：

$$\begin{aligned}\pi_{t+h} &= T_h(x_t) + u_{t+h}, \\ t &= 1, \dots, T, h = 1, \dots, H,\end{aligned}$$

其中 π_{t+h} 是未來的通貨膨脹率， t 為時間點， h 為預測期間； $x_t = (x_{1,t}, \dots, x_{k,t})'$ 為預測變數，共有 k 個變數，包含當期通貨膨脹率，以及各種可能影響通膨率的總體金融變數； u_{t+h} 是模型誤差項。本文的預測模型中， $T_h(\cdot)$ 可為線性或是非線性函數，也可以是單個模型或不同模型組合。給定 t 期(含)以前的資訊，藉由上述通膨率預測模型，我們可以得到 $t+h$ 期的通膨率直接預測(direct forecast)方程式：

$$\hat{\pi}_{t+h|t} = \hat{T}_{h,t-W+1,t}(x_t),$$

其中 $\hat{T}_{h,t-W+1,t}(\cdot)$ 是利用 $t-W+1$ 到 t 期的數據估計之目標函數， W 是預測視窗長度。本文參考過去文獻，考慮多種大數據預測模型，包含LASSO 模型家族、因子模型、非線性機器學習方法模型，以及組合式預測模型，並以傳統時間序列模型為基準模型。各種模型簡要介紹如下。

一、時間序列模型

為了提供模型比較基礎，我們將時間序列模型列為基準模型。本文考慮三個時間序列模型，第一個時間序列模型為隨機漫步模型，設定如下：

$$\pi_t = \pi_{t-1} + e_t;$$

第二個時間序列模型為自我迴歸 (AR(p)) 模型，設定如下：

$$\pi_t = \alpha_0 + \sum_{i=1}^p \alpha_i \pi_{t-i} + e_t;$$

第三個時間序列模型模型是自我迴歸與移動平均 (ARMA(p, q)) 模型，設定如下：

$$\pi_t = \alpha_0 + \sum_{i=1}^p \alpha_i \pi_{t-i} + \sum_{j=1}^q \gamma_j e_{t-j} + e_t,$$

以上模型的 π_t 為通膨率， α_i 、 γ_j 為模型參數， e_t 為誤差項， p, q 為前期通膨率及誤差項期數，本文使用貝氏資訊準則 (Bayesian information criterion; BIC) 選取最適的期數。

二、LASSO模型家族

本文考慮五種利用收縮法估計迴歸模型的 LASSO 模型家族，其中預測模型設定為線性迴歸模型，令 $y_t = \pi_{t+h}$ 代表迴歸模型

的被解釋變數， $T_h(x_t) = \beta_0 + \sum_{i=1}^k \beta_i x_{it}$ ， β_i 為參數， x_{it} 為第 i 個解釋變數。首先，第一個 LASSO 模型家族之模型為 Hoerl (1962) 和 Hoerl and Kennard (1970) 提出之脊迴歸方法，其極小化問題如下：

$$\min \sum_{t=1}^T (y_t - \beta_0 - \sum_{i=1}^k \beta_i x_{it})^2 + \lambda \sum_{i=1}^k \beta_i^2,$$

其中 T 為樣本數， λ 為調整係數，且 $\lambda \geq 0$ 。

第二個預測模型為 Tibshirani (1996) 提出的 LASSO 方法，其極小化的問題如下：

$$\min \sum_{t=1}^T (y_t - \beta_0 - \sum_{i=1}^k \beta_i x_{it})^2 + \lambda \sum_{i=1}^k |\beta_i|,$$

其中 λ 為調整係數，本文透過交叉驗證方式選取最適 λ 。再者，由 Zou and Hastie (2005) 所提出的 EN 方法目標式如下：

$$\min \sum_{t=1}^T (y_t - \beta_0 - \sum_{i=1}^k \beta_i x_{it})^2 + \lambda_1 \sum_{i=1}^k |\beta_i| + \lambda_2 \sum_{i=1}^k \beta_i^2,$$

其中 $\lambda_2/(\lambda_1 + \lambda_2) \geq 0$ ， λ_1 及 λ_2 均透過交叉驗證方式選取最適值。由上式可知，當 $\lambda_2 = 0$ 時，EN 方法可簡化為 LASSO 方法；當 $\lambda_1 = 0$ 時，EN 方法可簡化為脊迴歸。

由於 LASSO 方法在某些狀況下，所挑選的變數可能有不一致的問題，因此 Zou (2006) 提出 Adaptive LASSO，用以改進 LASSO 方法。Adaptive LASSO 方法之執行分為兩階段：第一階段時，估計具有一致性

(\sqrt{n} -consistency) 的估計式 $\tilde{\beta}_i$ ；在第二階段使用 $\hat{w}_i = 1/|\tilde{\beta}_i|^\gamma$ 為權重，其中 $\gamma > 0$ ，並求解下列極小化問題：

$$\min \sum_{t=1}^T (y_t - \beta_0 - \sum_{i=1}^k \beta_i x_{it})^2 + \lambda \sum_{i=1}^k \hat{w}_i |\beta_i|.$$

第五個 LASSO 模型家族為 Zou and Zhang (2009) 提出之 Adaptive EN 方法，其目標式如下，

$$\min \sum_{t=1}^T (y_t - \beta_0 - \sum_{i=1}^k \beta_i x_{it})^2 + \lambda_1 \sum_{i=1}^k w_i |\beta_i| + \lambda_2 \sum_{i=1}^k \beta_i^2,$$

其中 w_i 是參數 β_i 在 L_1 範數(norm)懲罰項中的權重，可使用與 Adaptive LASSO 法中同樣的方法決定， λ_1 及 λ_2 均透過交叉驗證方式選取最適值。

三、因子模型

因子模型法是從大量且高維度的變數中，萃取出因子變數，再利用因子變數進行預測。考慮 k 個預測變數，並設定以下因子模型：

$$x_{it} = \lambda_i' f_t + e_{it}, \quad i = 1, \dots, k, t = 1, \dots, T, \quad (1)$$

其中 $f_t = (f_t^1, \dots, f_t^r)'$ 代表這些變數的共同因子向量，維度為 $(r \times 1)$ ，且通常 $r \ll k$ ； λ_i 表示第 i 個變數的因子負擔(factor loading)，可視為因子向量 f_t 的權重， e_{it} 為代表這些變數個別特徵的誤差項。定義

$x_t = (x_{1t}, x_{2t}, \dots, x_{kt})'$, $e_t = (e_{1t}, e_{2t}, \dots, e_{kt})'$ 以及 $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)'$, 我們可以將因子模型(1)寫成矩陣形式：

$$x_t = \Lambda f_t + e_t.$$

若假設不同變數間的誤差項不相關，即 $E(e_{it}e_{jt}) = 0, \forall i \neq j$ ，此時因子模型為靜態模型，我們可利用主成分分析方法估計靜態因子模型，作法為利用特徵向量找出因子負擔矩陣 $\hat{\Lambda} = [\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_k]$ ，然後代入 $f_t = \Lambda' x_t$ 可得因子估計式： $\hat{f}_t = \hat{\Lambda}' x_t$ 。

Bai and Ng (2008)認為，若直接由所有蒐集到的預測變數矩陣萃取因子，其中無資訊變數會添加許多雜訊到因子的估計中，使得因子的預測能力下降。他們建議排除無資訊變數，選擇有預測能力的目標預測變數 (targeted predictors)，然後再估計因子，此即為目標因子方法。Bai and Ng (2008)建議可用硬門檻或軟門檻標準選擇變數，本文使用硬門檻標準選擇目標預測變數，接著利用因子模型方法估計及選出適當的因子變數 $\hat{f}_t = (\hat{f}_t^1, \dots, \hat{f}_t^r)'$ ，再將這些因子變數加入自我迴歸模型，建構預測模型。此作法為 Stock and Watson (2002)的擴散指標預測方程式：

$$\pi_{t+h} = \alpha_0 + \sum_{i=1}^p \alpha_i \pi_{t+h-i} + \sum_{j=0}^q \gamma_j \hat{f}_{t-j} + e_{t+h},$$

通常以BIC選擇最適落後期數 p, q 。本文考慮兩種因子模型方法，分別是靜態因子模

型與目標因子模型。為簡化起見，我們考慮不含落後項的預測模型，故在估計出因子 \hat{f}_t 後，將 \hat{f}_t 視為可觀察變數，因此向前 h 期預測為：

$$\hat{\pi}_{t+h} = \hat{\alpha}_0 + \hat{\gamma}_0 \hat{f}_t.$$

四、非線性機器學習方法

晚近以決策樹(decision tree)為基礎之機器學習法常被用來預測通膨，這些方法最大的特點在於，透過決策樹可以有效地捕捉變數間的非線性關係。在建構決策樹模型時，由Breiman et al.(1984)所提出的 CART (classification and regression tree)是常用的建構決策樹方法。本文考慮兩種非線性機器學習方法，首先是由Breiman (1996)所提出的套袋法，我們利用區塊拔靴抽樣法(block bootstrap)生成多組樣本，並訓練決策樹，再將這些決策樹之預測結果平均，以降低決策樹訓練時的變異。在本文中，我們以全部解釋變數為預測變數，故每組新樣本集合有 k 個預測變數，通膨率為被預測變數，接著建構 B 棵決策樹，每一棵決策樹會有一個對應的預測值，簡單平均這 B 個預測值即可得到套袋樹方法的預測值。

本文考慮的第二種非線性機器學習方法為隨機森林方法，隨機森林為Breiman(2001)所提出，與套袋樹方法類似，隨機森林建構每一決策樹的資料都由原訓練資料以拔靴法抽出，之後再將這些決策樹之預測值平

均。但和套袋樹法不同的是，在訓練時，每個決策樹之節點在分岔時會隨機從 k 個變數當中選 p 個計算類別極端化程度，再用其中表現最好的變數建構分岔。選取之變數數目可依預測模型而有所不同，如迴歸模型常用 $[k/3]$ 個變數，分類模型常用 $[\sqrt{k}]$ 個變數，其中 k 為全部變數個數， $[\]$ 為高斯符號。由於在建構決策樹的過程加入了隨機性，因此可以解決過度配適的問題。又隨機選取變數的方式可以降低拔靴法訓練的決策樹之間的相关性，進而降低決策樹訓練的變異數。使用隨機森林方法預測通膨率之實際作法與上述之套袋樹作法類似，唯一不同之處為，每一個拔靴樣本建構決策樹時，在決策樹每一個節點，隨機挑選 $[k/3]$ 個解釋變數建構分岔。

五、組合式預測方法

眾多預測方法中，組合式預測方法也常被使用。組合式預測方法的概念為組合不同

預測，以求取預測表現最佳之方法。本文考慮四種組合式預測方法，前兩種組合式預測方法為利用 LASSO 模型家族選出變數後，再利用所選變數做為模型預測變數，也就是分別以 LASSO 或是 EN 方法挑選變數，再依挑選之變數設定線性迴歸模型，並以此預測通膨率。第三種組合式預測方法為 LASSO-套袋樹方法，此組合方法先使用 LASSO 進行變數挑選，然後使用 LASSO 所挑選的變數建構新資料，並以套袋樹法建構決策樹並進行預測。若此混合方法與單純使用套袋樹法的結果相近的話，表示是否挑選變數並不重要。第四種組合方法為 Medeiros et al. (2021) 所提出的混合線性隨機森林 (hybrid linear-random forest models)，此方法結合非線性的套袋樹與線性迴歸方法，先利用建構決策樹時選取變數，並以所選變數建構線性迴歸模型。由於篇幅限制，詳細的估計及預測方法，以及參數設定等，敬請參考央行委託研究計畫完整報告。

參、資料說明與模型比較

一、台灣通貨膨脹率

本文通膨率定義如下：

$$\pi_t = \frac{CPI_t - CPI_{t-12}}{CPI_{t-12}} \times 100,$$

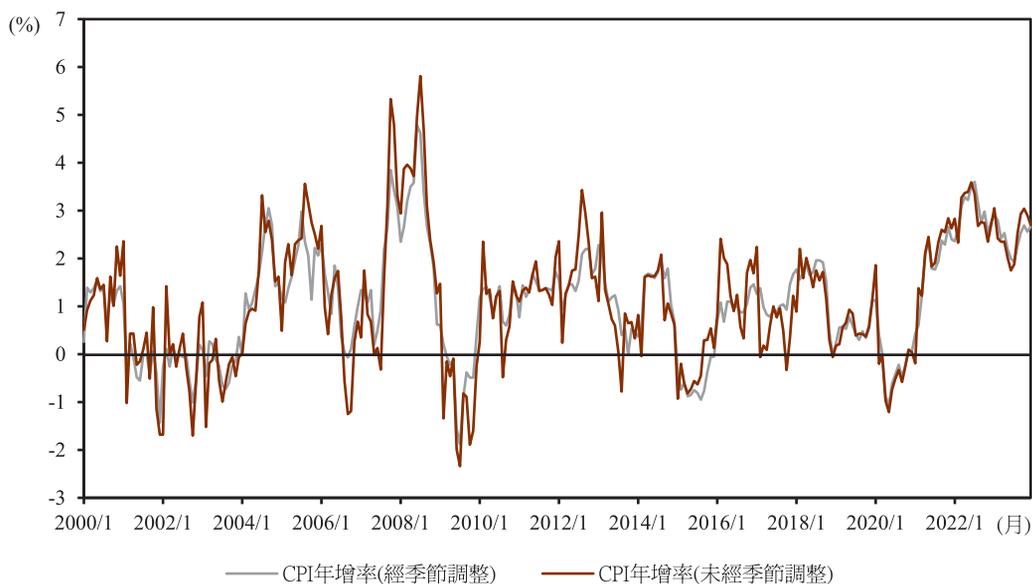
其中， CPI_t 為 t 時點(月)季調後CPI，因此通貨膨脹率 π_t 即為季調後CPI年增率。圖1為

2000年1月至2023年12月間，季調前後的台灣通膨率走勢。我們可將台灣通膨率走勢大致分為三個階段，第一階段為2000年至2010年。在2003年以前，通膨率在零軸上下2個百分點之間波動，自2004年起逐漸走升，至2006年緩步回降。又自2007年下半年起，受

到全球金融危機影響，國際原物料及原油價格持續大幅上揚，致使台灣通膨率明顯上漲，此一漲勢至2009年回跌。第二階段為2010年至2020年，此一階段台灣通膨率走勢相對穩定，維持較低且穩定水準。第三階段

為2021年迄今，受到COVID-19疫情導致全球供應鏈瓶頸及俄烏戰爭影響，國際原物料及原油價格再度攀升，推升食物類、油料費及耐久性消費品價格，通膨率再度走升。

圖1 台灣通膨率走勢

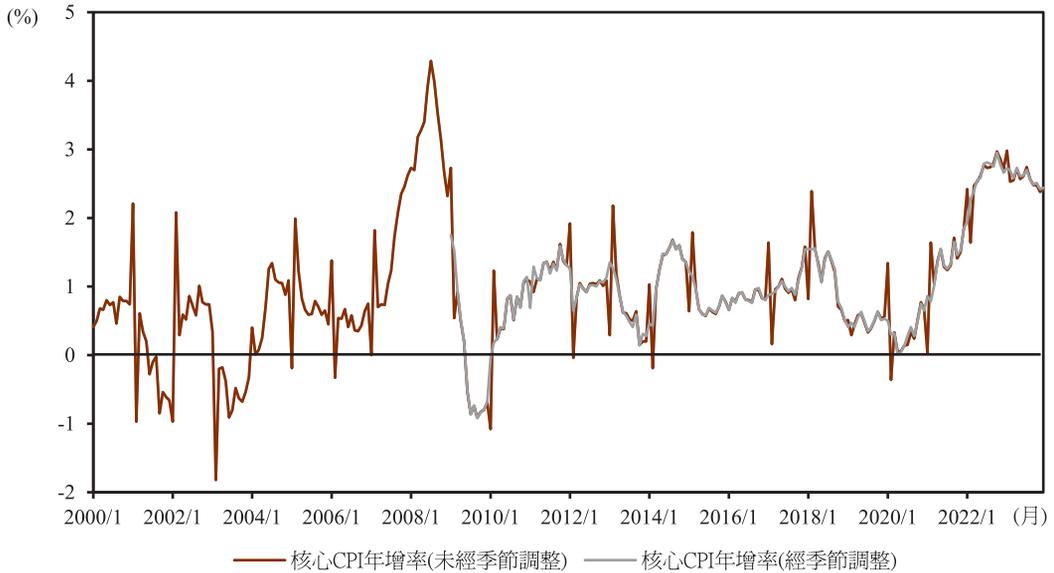


資料來源：中華民國統計資訊網。

圖2為相同期間下，排除蔬果類與能源類之核心CPI年增率計算之台灣核心通膨率走勢。一般而言，核心CPI已排除波動較大

的干擾因素，因此核心通膨率的漲跌變化與一般通膨率大致相同，惟核心通膨率的變動幅度較小。

圖2 台灣核心通膨率



資料來源：中華民國統計資訊網。

註：經季調之核心CPI通膨率最早的資料期間為2009年。

二、大數據資料

本文選擇93個與台灣通貨膨脹息息相關的總體經濟及貨幣金融變數，並將這些變數分成11個類別，包括：景氣指標、就業、民生消費、貨幣金融、股票市場、債券市場、房地產市場、大宗物資、國際收支與進出口、海外因素及情緒指標。以下就各個類別所含變數進行說明。

景氣指標方面，本文收錄台灣實質GDP年增率、景氣對策信號綜合分數、領先指標綜合指數、同時指標綜合指數、落後指標綜合指數、落後指標/領先指標、痛苦指數、工業生產指數年增率，以及工業生產指數-製造業年增率等實質面與名目面的景氣相關指標。資料來源為台灣經濟新報(Taiwan

Economic Journal, TEJ)與經濟部統計處。就業方面，根據Stock and Watson (1999)研究，失業率有助於預測通貨膨脹率，因此本文將失業率納入預測變數。除此之外，本文嘗試納入勞動參與率、非農業就業人口、經常性薪資、基本工資-月、各行業平均工時-月，以及工業及服務業受僱員工淨進入率等就業相關變數，觀察這些變數是否對提升通膨預測準確性有所助益。資料來源為TEJ與中華民國統計資訊網。

民生消費方面，本文將季調CPI年增率及未季調CPI年增率均納入預測變數，並納入未季調核心CPI年增率、未季調躉售物價總指數年增率、進口物價總指數、出口物價總指數、批發業營業額年增率，以及零售業

營業額年增率等。值得注意的是，行政院主計總處自2023年1月起停編躉售物價指數，故後續資料使用「國產與進口品物價總指數」所計算之年增率進行替代。資料來源為中華民國統計資訊網與經濟部統計處。貨幣金融方面，考量貨幣總計數對通膨有直接影響，本文將M1B年增率與M2年增率均納入預測變數。此外，本文亦將金融機構存放款情形相關變數納入預測變數，如信用對產出佔比、本國銀行逾放比率、台灣全體貨幣機構放款、台灣全體貨幣機構存款、台灣全體貨幣機構存放比，以及台灣全體金融機構資產淨額等。資料來源為TEJ與中央銀行，信用產出比與台灣全體貨幣機構存放比則係自行計算而得。

除上述常見的總體經濟指標外，本文同時考慮股票市場、債券市場，以及房地產市場等金融市場重要變數可能蘊含的通膨預測資訊。本文選擇的股票市場變數包含台股加權指數、金融類股票指數、機電類股票指數、大盤週轉率、大盤股價淨值比、大盤本益比、大盤股價營收比、大盤現金殖利率、恐慌指數，以及Ami非流動性指數。資料來源皆為TEJ，而Ami非流動性指數為參考Amihud (2002)與Barardehi et al. (2021)，計算方式請參考央行研究計畫之完整報告。另，許多重要的債券市場變數與通膨密切相關。根據預期理論，長天期公債利率隱含一國未來景氣與通膨狀況。Sims (1980)研究發現，

商業本票利率有助於預測通膨；而Mishkin (1989)、Jorion and Mishkin (1991)等研究則發現，長短天期利差對預測通膨亦有所助益。本文因此考慮重貼現率、金融業拆款利率、五年期公司債、台灣十年期公債利率、商業本票31-90天期次級市場利率、美國長(十年)短(五年)天期利差，以及美國長(十年)短(一年)天期利差等變數作為預測變數，期望借助這些變數與通膨有關的資訊內涵，改善通膨預測表現。資料來源為TEJ與中央銀行，利差相關變數係自行計算而得。

鑑於CPI分類項目中，居住類價格的權重僅次於食物類價格，對通膨具一定程度影響，加以Stock and Watson (1999)發現消費者購屋貸款以及生產者建屋貸款等變數對於通膨有良好的預測能力，因此本文亦納入台北市房租指數、五大銀行新承做放款金額-購屋貸款、五大銀行新承做放款利率-購屋貸款、消費者購屋貸款(購置住宅貸款與房屋修繕貸款)、生產者建屋貸款(建築貸款)、消費者購屋貸款集中度，以及生產者建屋貸款集中度等房地產市場相關變數。資料來源為中央銀行、中華民國統計資訊網，以及台北市房租指數銜接表，其中購屋貸款與建屋貸款集中度係自行計算而得。

由於台灣為一小型開放經濟體，國內經濟狀況深受國際經濟情勢影響，故本文亦納入國際收支與進出口相關變數，包含金融帳-資產-其他投資、金融帳-負債-其他投

資、經常帳淨額、外匯存底、外銷訂單、出(入)超、價值指數-出口，以及價值指數-進口等。資料來源為TEJ與財政部。另外，台灣具出口導向經濟體的特性，國際原物料及原油價格等供給面因素對通膨影響甚鉅，因此本文納入銅現貨、黃金現貨、鋁現貨、銅板、布蘭特原油現貨，以及CRB期貨指數等變數，其中，CRB期貨指數包含19項商品，被用來衡量國際商品價格走勢。上述變數資料來源皆為TEJ。

除國際原物料及原油價格外，本文針對中國大陸、美國、日本及南韓4個主要貿易往來國家，納入可能影響通膨的因素，包含：中國實質GDP成長率、中國CPI年增率、日本實質GDP成長率、日本十年期公債殖利率、日本CPI年增率、日本隔夜無擔保拆借利率、美國實質GDP成長率、美國十年期公債殖利率、美國二年期公債殖利率、美國三個月期國庫券利率、美國CPI年增率、美國聯邦資金市場利率、南韓實質GDP成長率、南韓公債殖利率、南韓CPI年增率、南韓貨幣市場利率、美日匯率、美中匯率、美韓匯率、美台匯率。除了美國實質GDP成長率資料取自聖路易聯邦準備銀行資料庫(Federal Reserve Economic Data, FRED)外，其餘資料來源皆為TEJ。

最後，本文考量經濟個體(廠商或家戶)根據其對於未來通膨或景氣的態度，調整經濟決策，進而對整體經濟情況有重大影響。

鑑此，本文收錄情緒指標，探討這些指標對於通膨的預測能力。情緒指標可進一步區分為景氣面、企業面及股市面。我們使用國立新加坡大學信用風險評等系統所計算的市值平均企業脆弱指數，作為企業面之情緒衡量標準；並使用中央大學台灣經濟發展研究中心所編製的消費者信心指數作為景氣面之情緒衡量標準。至於股市面之情緒衡量標準，我們參考Tobias and Brunnermeier (2016)計算Delta CoVaR，該變數用於衡量在金融機構不同風險事件的條件下，大盤月報酬率的風險價值。本文所採用之許多變數，其歷史資料可能面臨持續更新或調整，本文以2023年12月31日之資料為基準，其中，資料的處理依照以下三個步驟進行：(1)調整所有變數的資料頻率一致；(2)處理資料遺漏值；(3)進行單根檢定及定態調整。

三、模型比較

本文以3個時間序列模型、5個LASSO模型家族、2個因子模型、2個非線性機器學習模型及4個組合模型預測未來通膨率 π_{t+h} ，並使用滾動視窗法(rolling window)評估模型的樣本外預測能力。假設模型估計所使用之樣本數，即視窗長度為 W 。以在第 t 期預測第 $t+h$ 期之通膨率 π_{t+h} 來說，我們使用第 $t-W+1$ 至第 t 期的樣本估計並產生預測值 $\hat{\pi}_{t+h}$ 。本文設定之 $W=120$ ， $h=(1, 3, 6, 12, 24)$ ，起始訓練樣本期間

為2000年3月至2010年2月，樣本外期間為2010年3月至2023年12月，以向前一期預測 ($h = 1$) 為例，共產生166個預測點，依此類推。

我們以均方根誤差 (root mean squared error, RMSE)、平均絕對誤差(mean absolute error, MAE)，以及中位數絕對偏差(median absolute deviation, MAD)等指標評估模型的樣本外預測表現。RMSE 為衡量預測值與實際觀測值之間差異的平均幅度的指標，計算方式為誤差平方的平均之平方根：

$$RMSE = \sqrt{\frac{1}{T - (t + h) + 1} \sum_{i=t+h}^T \hat{e}_i^2},$$

其中 $\hat{e}_i = \pi_i - \hat{\pi}_i$ 為預測誤差， π_i 是實際觀測值， $\hat{\pi}_i$ 是樣本外預測值。預測

值由第 $t + h$ 期開始至第 T 期，總共有 $T - (t + h) + 1$ 個樣本外預測。

由於 RMSE 易受極端值影響，許多研究也考慮使用 MAE 為評估預測表現的指標，計算方式為誤差絕對值的平均：

$$MAE = \frac{1}{T - (t + h) + 1} \sum_{i=t+h}^T |\hat{e}_i|.$$

MAD 則是藉由計算所有誤差與誤差之中位數之差異的絕對值後，對這些絕對值取中位數：

$$MAD = \text{median}(|\hat{e}_i - \text{median}(\{\hat{e}_{t+h}, \dots, \hat{e}_T\})|), i = t + h, \dots, T.$$

由上所述，以單一指標來看，由模型預測誤差所計算之指標數值越低，表示模型預測能力較佳，預測表現越好，但由於計算標準不同，故不同指標之間無法相互比較。

肆、實證結果與討論

本文應用總體經濟及金融市場大數據資料，以3個時間序列模型，和13個機器學習模型預測台灣通膨率，並討論預測模型表現。此外，透過其中部分預測模型挑選變數，我們得以識別隱含重要通膨預測資訊的預測變數。以下分別以向前一期通膨預測、一年以內的短期通膨預測，以及二年的中長期通膨預測評估模型。

一、向前一期通膨預測

本文首先進行 16 個預測模型的樣本外預測結果評估，向前一期的預測表現列於表1，不同模型下預測誤差最小值以粗體呈現。由表1可知，以RMSE為模型表現評估標準時，時間序列模型(基準模型)中以ARMA模型表現最佳，其RMSE為0.509；LASSO模型家族以Adaptive LASSO模型表現最佳，RMSE為0.372，預測誤差最小，明顯優於時間序列模型的預測表現；因子模型之RMSE為所有預測模型中最大，表示模型

的預測表現不佳。而不論是機器學習模型或是組合模型，其向前一期之預測表現差異不大，RMSE介於0.381與0.468之間。以MAE評估模型表現來看，LASSO模型家族的Adaptive LASSO模型預測表現最佳，MAE為0.290，其次為Adaptive EN及LASSO模型。結合LASSO模型的LASSO-套袋樹模型或是LASSO-LS模型的預測亦表現良好。而若以MAD評估模型表現，則是以組合模型表現最佳，LASSO-LS模型及LASSO-套袋樹模型的MAD均為0.237，LASSO模型家族及非線性機器學習模型預測表現居中，接著是時間序列模型，最後是因子模型。

綜上所述，觀察實證結果可以發現，LASSO模型家族的預測表現普遍較基準模型佳，以RMSE及MAE作為評估指標時，使用Adaptive LASSO模型的表現最佳；而以MAD作為評估指標時，使用LASSO-LS及LASSO-套袋樹模型的預測表現最佳。就因子模型而言，此類模型的預測表現較基準模型差，若以目標因子模型進行預測，雖可改善一般因子模型(主成分分析法)的預測表現，惟改善幅度有限，其表現仍較基準之時間序列模型差。就機器學習模型及組合模型而言，實證結果顯示，這些模型的預測表現亦較基準模型佳，與LASSO模型家族的預測表現差異不大，而非線性機器學習模型中，以套袋樹法表現較佳，組合模型中則以LASSO-LS及LASSO-套袋樹法表現較佳。

圖3為向前一期通膨預測的樣本外預測走勢圖，其中黑色實線為樣本外真實值。由這些走勢圖可以發現，組合模型樣本外預測走勢與真實值亦步亦趨，能夠精確地預測台灣通膨率。LASSO模型家族與非線性機器學習模型之預測走勢也相當接近真實值，除當通膨率落於谷底或高峰等較不易被捕捉外，其樣本外預測能力也相當不錯。以模型預測走勢與真實值之相近程度而言，上述三類模型之樣本外預測值優於基準時間序列模型之表現；而因子模型明顯表現較差，其預測走勢與真實值有一段落差。

最後，本文列出使用LASSO模型家族、因子模型、非線性機器學習模型及組合模型選擇之重要變數於表2。就LASSO模型家族來看，我國前一年經季調的CPI年增率、中國CPI年增率、中國實質GDP成長率、美國CPI年增率、布蘭特原油現貨、出口物價總指數、M2年增率、批發業營業額年增率，以及零售業營業額年增率等變數為四個LASSO模型家族模型大多選擇的變數。此結果與台灣屬小型開放經濟體，深受國際經濟情勢與進出口貿易影響的直覺相符，而批發業與零售業營業額等重要內需市場指標，以及貨幣總計數等項目，均對通膨有重要影響。

就因子模型而言，我們以目標因子模型進行變數挑選，該模型選擇CPI年增率、批發業營業額年增率、零售業營業額年增率，

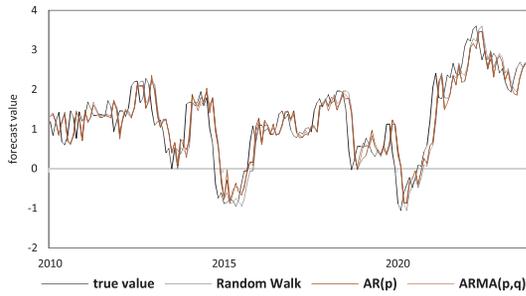
以及出口物價總指數與LASSO模型家族相同外，另選取金融類股票指數、台灣全體金融機構資產淨額，以及消費者購屋貸款集中度等變數。就非線性機器學習模型及組合模型而言，這些模型均選擇前一年經季調的CPI年增率、中國CPI年增率、美國CPI年增率、躉售物價總指數年增率，以及非流動性

指標。由此可以發現，除部分與LASSO模型家族重複選取的重要變數外，此類模型特別篩選躉售物價總指數年增率為預測通膨重要變數，推測係因躉售物價指數主要衡量原物料與零件等中間財價格，其變動反映採購成本的變化，因此富含通膨預測資訊。

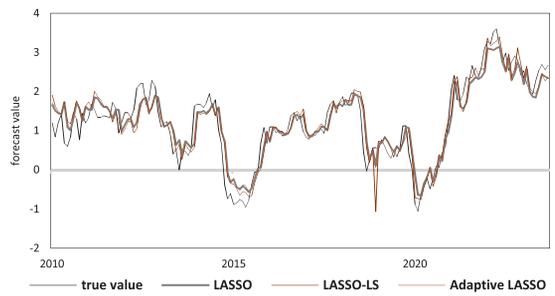
表1 向前一期預測表現($h = 1$)

預測誤差計算方式			
模型	RMSE	MAE	MAD
時間序列模型			
隨機漫步	0.525	0.393	0.290
AR(p)	0.511	0.392	0.301
ARMA(p,q)	0.509	0.387	0.305
LASSO模型家族			
LASSO	0.381	0.296	0.259
脊迴歸	0.525	0.403	0.322
EN	0.381	0.298	0.263
Adaptive LASSO	0.372	0.290	0.246
Adaptive EN	0.378	0.295	0.254
因子模型			
主成分分析法	1.010	0.837	0.741
目標因子	0.761	0.523	0.321
非線性機器學習模型			
套袋樹法	0.450	0.345	0.261
隨機森林	0.468	0.366	0.283
組合模型			
LASSO-LS	0.396	0.303	0.237
EN-LS	0.399	0.305	0.262
LASSO-套袋樹	0.381	0.294	0.237
隨機森林-LS	0.414	0.318	0.268

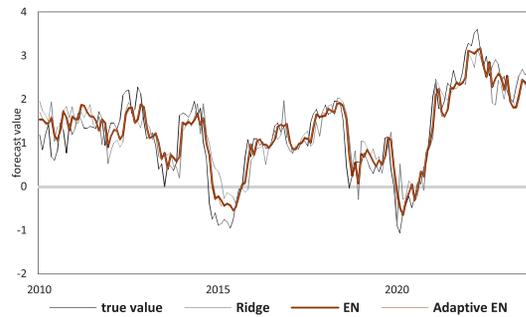
圖3 向前一期樣本外預測走勢圖



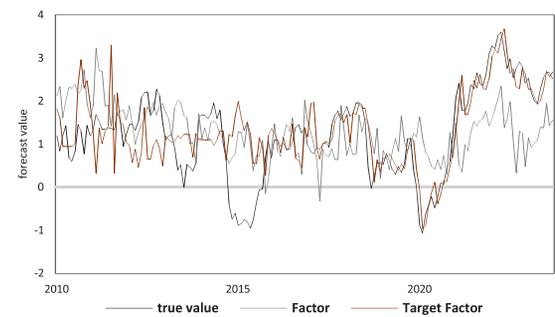
(1) 基準模型



(2) LASSO 模型家族 1



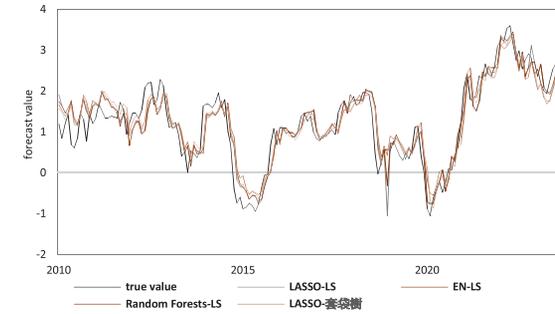
(3) LASSO 模型家族 2



(4) 因子模型



(5) 非線性機器學習模型



(6) 組合模型

註：LASSO模型家族1除包括原有LASSO模型家族中之LASSO模型外，我們加入LASSO-LS模型之預測走勢，以便於比較及參考。

表2 向前一期各模型重要變數

排序	LASSO	EN	Adaptive LASSO	Adaptive EN	目標因子
1	CPI年增率(經季調)	CPI年增率(經季調)	CPI年增率(經季調)	CPI年增率(經季調)	CPI年增率(經季調)
2	中國CPI年增率	中國CPI年增率	批發業營業額年增率	中國CPI年增率	金融類股票指數
3	批發業營業額年增率	批發業營業額年增率	中國CPI年增率	批發業營業額年增率	消費者購屋貸款集中度
4	零售業營業額年增率	零售業營業額年增率	零售業營業額年增率	痛苦指數	出(入)超
5	中國實質GDP成長率	痛苦指數	中國實質GDP成長率	零售業營業額年增率	台灣全體金融機構資產淨額
6	出口物價總指數	美國CPI年增率	出口物價總指數	美國CPI年增	零售業營業額年增率
7	布蘭特原油現貨	M2年增率	M2年增率	中國實質GDP成長率	批發業營業額年增率
8	美國CPI年增率	中國實質GDP成長率	布蘭特原油現貨價格	M2年增率	出口物價總指數
排序	套袋樹	隨機森林	LASSO-套袋樹	隨機森林-LS	
1	CPI年增率(經季調)	CPI年增率(經季調)	CPI年增率(經季調)	CPI年增率(經季調)	
2	美國CPI年增率	美國CPI年增率	中國CPI年增率	痛苦指數	
3	中國CPI年增率	臺灣物價總指數年增率(未經季調)	臺灣物價總指數年增率(未經季調)	臺灣物價總指數年增率(未經季調)	
4	臺灣物價總指數年增率(未經季調)	中國CPI年增率	美國CPI年增率	美國CPI年增率	
5	非流動性指標	痛苦指數	M2年增率	非流動性指標	
6	M2年增率	非流動性指標	痛苦指數	中國CPI年增率	
7	進口物價總指數	M2年增率	大盤股價營收比	進口物價總指數	
8	台灣十年期公債利率	重貼現率	非流動性指標	台灣十年期公債利率	

二、短期通膨預測

本文期亦利用各種預測模型進行一年內之短期通膨率預測，本小節討論這些模型對未來3、6、12個月的通膨率預測表現。首先，表3列示向前3個月及6個月通膨預測表現，不同模型下預測誤差最小值以粗體呈現。相較於向前一期之預測表現，向前3或6期之整體預測誤差皆有上升趨勢。實證結果顯示，無論使用何種評估標準，非線性機器學習模型的預測誤差皆最小。LASSO 模型家族或是時間序列模型在向前3期之預測表現接近，但向前6期之預測表現以LASSO 模型家族較佳。整體而言，在預測未來3個月通膨時，以RMSE、MAE及MAD衡量表現最佳的模型分別為套袋樹法、隨機森林及AR模型，其預測誤差值分別為0.631、0.509及0.411；在預測未來6個月通膨時，不論是以何種誤差指標衡量，均為隨機森林的預測表現最佳，其RMSE、MAE及MAD值分別為0.659、0.546及0.472。向前12個月通膨預測表現，則可由表4觀之。相較於基準模型，

其他模型均有助於提升預測準確性，其中以非線性機器學習模型表現最佳，LASSO模型家族次之，並略優於組合模型，而因子模型表現最差。整體而言，在預測未來12個月通膨時，不論是以何種誤差指標衡量，均為非線性機器學習模型中的隨機森林方法預測表現最佳。

不同預測期間之下，各預測模型所選擇的重要變數各有不同。以表5為例，在向前3個月通膨預測時，LASSO模型家族的變數選擇大致相同，除以前一年經季調的CPI年增率及M2年增率等與預測目標直接相關的變數外，大多選擇中國CPI年增率、中國實質GDP成長率、美國長短天期利差、美國CPI年增率、美台匯率、南韓CPI年增率，以及進口物價總指數等，顯示台灣通膨變動容易受到貿易對手國經濟因素影響。目標因子模型、非線性機器學習模型及組合模型除亦選擇許多國際因素相關變數外，均選擇躉售物價總指數年增率，這說明供給面因素對通膨預測的重要性。

表3 未來一年內預測表現 ($h = 3、6$)

模型	RMSE		MAE		MAD	
	h=3	h=6	h=3	h=6	h=3	h=6
時間序列模型						
隨機漫步	0.745	1.006	0.551	0.789	0.420	0.680
AR(p)	0.712	0.960	0.527	0.755	0.411	0.660
ARMA(p,q)	0.715	0.906	0.539	0.713	0.439	0.534
LASSO模型家族						
LASSO	0.730	0.756	0.579	0.640	0.471	0.572
脊迴歸	0.709	0.834	0.568	0.689	0.483	0.627
EN	0.713	0.776	0.566	0.637	0.466	0.594
Adaptive LASSO	0.720	0.769	0.574	0.649	0.481	0.597
Adaptive EN	0.691	0.786	0.554	0.651	0.449	0.585
因子模型						
主成分分析法	1.019	1.077	0.850	0.888	0.854	0.832
目標因子	0.985	0.980	0.796	0.797	0.714	0.676
非線性機器學習模型						
套袋樹法	0.631	0.674	0.513	0.563	0.428	0.491
隨機森林	0.633	0.659	0.509	0.546	0.423	0.472
組合模型						
LASSO-LS	0.811	0.740	0.630	0.610	0.510	0.563
EN-LS	0.792	0.749	0.608	0.610	0.504	0.545
LASSO-套袋樹	0.686	0.808	0.555	0.679	0.440	0.611
隨機森林-LS	0.710	0.794	0.577	0.665	0.490	0.583

在向前6個月通膨預測方面（見表6），重要預測變數有相當程度的改變。LASSO模型家族除選擇一定比例的國際因素相關變數，如中國CPI年增率、美國三個月期國庫券利率等，亦選擇零售業營業額年增率及本國銀行逾放比率等國內市場變數。而因子模型、非線性機器學習模型及組合模型方面，除選擇一定比例的國際因素相關變數外，也

同時選出領先、同時、落後指標綜合指數等景氣相關變數。值得注意的是，在預測未來6個月通膨時，許多預測模型選擇之重要變數包含股票市場大盤相關指標，以及台北市房租指數與五大銀行新承做放款利率-購屋貸款等變數，顯示金融市場及不動產市場中亦隱含預測短期通膨重要資訊。

表4 未來一年及二年預測表現 ($h = 12, 24$)

模型	RMSE		MAE		MAD	
	h=12	h=24	h=12	h=24	h=12	h=24
時間序列模型						
隨機漫步	1.287	1.551	1.029	1.224	0.855	1.025
AR(p)	1.245	1.491	1.003	1.182	0.807	1.012
ARMA(p,q)	1.165	1.362	0.943	1.094	0.711	0.901
LASSO模型家族						
LASSO	0.821	0.779	0.643	0.637	0.516	0.536
脊迴歸	0.815	0.860	0.632	0.710	0.524	0.622
EN	0.857	0.816	0.628	0.667	0.460	0.570
Adaptive LASSO	0.810	0.789	0.615	0.640	0.470	0.564
Adaptive EN	0.816	0.841	0.612	0.698	0.473	0.625
因子模型						
主成分分析法	0.966	1.072	0.745	0.850	0.519	0.657
目標因子	0.899	1.013	0.666	0.805	0.491	0.642
非線性機器學習模型						
套袋樹法	0.712	0.677	0.544	0.530	0.425	0.380
隨機森林	0.677	0.667	0.527	0.526	0.384	0.405
組合模型						
LASSO-LS	0.855	0.893	0.645	0.721	0.477	0.613
EN-LS	0.941	0.983	0.691	0.775	0.502	0.687
LASSO-套袋樹	0.897	0.849	0.711	0.729	0.601	0.552
隨機森林-LS	0.887	0.817	0.706	0.695	0.551	0.519

在12個月通膨預測方面(見表7)，LASSO模型家族選擇之重要變數除前一年經季調的CPI年增率外，仍以國際因素相關變數為主，包含美國聯邦資金市場利率、美國、中國及日本CPI年增率，以及CRB期貨指數。因子模型、非線性機器學習模型及組合模型方面，日本CPI年增率的重要性有所增加，再者，除與先前相同的領先、落後指

標綜合指數等景氣指標外，重要變數新增商業本票31-90天期次級市場利率、金融業拆款利率等貨幣市場重要指標。最後，由表7可觀察到，大部分模型均選擇五大銀行新承做放款利率-購屋貸款為預測未來12個月通膨時重要變數，且重要程度與極短期預測相比有上升趨勢。

由上述結果可知，在預測未來一年內通

膨時，LASSO模型家族篩選相當數量的國際因素或進出口貿易相關變數，如美國、中國及日本 CPI 年增率，以及美台、美中、美日匯率等。在極短期預測時($h = 1, 3$)，除選擇前一年的 CPI 年增率為重要變數外，LASSO模型家族亦選擇進、出口物價總指數、零售業與批發業營業額年增率為重要變數。在短期預測時($h = 6, 12$)，LASSO模型家族選出較多國際及國內相關變數，國際變數包括美國三個月期國庫券利率、日本實質 GDP 成長率及美國聯邦資金市場利率，國內變數則包括台北市房租指數、五大銀行新承做放款利率-購屋貸款，以及本國銀行逾放比率等。非線性機器學習模型及組合模型亦篩選國際因素相關變數，如美國及中國 CPI 年增率。在極短期預測時($h = 1, 3$)，非線性機器學習模型之重要變數包括前一年的 CPI 年增率、躉售物價總指數年增率及 M2 年增率；在短期預測時($h = 6, 12$)的重要變數則包括國際變數，如中國及日本實質 GDP 成

長率，國內變數則如領先、同時、落後指標綜合指數等景氣指標及五大銀行新承做放款利率-購屋貸款、商業本票31-90天期次級市場利率、金融業拆款利率等重要市場利率變數。

三、中長期通膨預測

除短期通膨率外，本文亦嘗試運用大數據資料於中長期台灣通膨率預測。我們以向前24個月(未來二年)的通膨預測為例。由表4觀之，仍以非線性機器學習模型表現最佳，LASSO模型家族次之，並略優於組合模型，而因子模型表現最差，此一結果與短期通膨預測中，預測未來12個月通膨表現的結果相似。整體而言，在預測未來24個月通膨時，不論是以何種誤差標準衡量，隨機森林的預測表現最佳，以RMSE衡量，預測誤差值為0.667；而套袋樹法的預測表現亦相當良好，其MAD值為0.380，表現較隨機森林的0.405更佳。

表5 各模型重要變數($h = 3$)

排序	LASSO	EN	Adaptive LASSO	Adaptive EN	目標因子
1	CPI年增率(經季調)	CPI年增率(經季調)	CPI年增率(經季調)	CPI年增率(經季調)	中國CPI年增率
2	中國CPI年增率	中國CPI年增率	中國CPI年增率	中國CPI年增率	台灣十年期公債利率
3	美國長(十年)短(五年)天期利差	美國長(十年)短(五年)天期利差	美國長(十年)短(五年)天期利差	美國長(十年)短(五年)天期利差	同時指標綜合指數
4	批發業營業額年增率	批發業營業額年增率	批發業營業額年增率	批發業營業額年增率	南韓實質GDP成長率
5	進口物價總指數	進口物價總指數	進口物價總指數	進口物價總指數	躉售物價總指數年增率(未經季調)
6	美國CPI年增率	美國CPI年增率	美國CPI年增率	美國CPI年增率	台灣全體貨幣結構放款
7	南韓CPI年增率	美台匯率	南韓CPI年增率	美台匯率	銅現貨價格
8	美台匯率	南韓CPI年增率	美台匯率	M2年增率	中國實質GDP成長率
排序	套袋樹	隨機森林	LASSO-套袋樹	隨機森林-LS	
1	中國CPI年增率	中國CPI年增率	CPI年增率(經季調)	躉售物價總指數年增率(未經季調)	
2	躉售物價總指數年增率(未經季調)	躉售物價總指數年增率(未經季調)	中國CPI年增率	非流動性指標	
3	台灣十年期公債利率	台灣十年期公債利率	美國CPI年增率	台灣十年期公債利率	
4	非流動性指標	非流動性指標	痛苦指數	中國CPI年增率	
5	落後指標綜合指數	美國CPI年增率	M2年增率	同時指標綜合指數	
6	同時指標綜合指數	同時指標綜合指數	躉售物價總指數年增率(未經季調)	痛苦指數	
7	美國CPI年增率	CPI年增率(經季調)	批發業營業額年增率	美國CPI年增率	
8	五大銀行新承做放款利率-購屋貸款	金融業拆款利率	大盤股價營收比	落後指標綜合指數	

表6 各模型重要變數(h = 6)

排序	LASSO	EN	Adaptive LASSO	Adaptive EN	目標因子
1	中國CPI年增率	零售業營業額年增率	中國CPI年增率	中國CPI年增率	日本實質GDP成長率
2	台北市房租指數	中國CPI年增率	零售業營業額年增率	零售業營業額年增率	同時指標綜合指數
3	零售業營業額年增率	台北市房租指數	台北市房租指數	台北市房租指數	本國銀行逾放比率
4	美日匯率	美日匯率	美日匯率	美日匯率	消費者購屋貸款集中度
5	本國銀行逾放比率	本國銀行逾放比率	本國銀行逾放比率	本國銀行逾放比率	五大銀行新承做放款利率-購屋貸款
6	CRB期貨指數	CRB期貨指數	CRB期貨指數	美國三個月期國庫券利率	美國實質GDP成長率
7	痛苦指數	美國三個月期國庫券利率	美國三個月期國庫券利率	CRB期貨指數	MIB年增率
8	美國三個月期國庫券利率	痛苦指數	痛苦指數	痛苦指數	M2年增率
排序	套袋樹	隨機森林	LASSO-套袋樹	隨機森林-LS	
1	中國實質GDP成長率	同時指標綜合指數	中國CPI年增率	領先指標綜合指數	
2	領先指標綜合指數	領先指標綜合指數	大盤股價淨值比	五大銀行新承做放款利率-購屋貸款	
3	同時指標綜合指數	中國實質GDP成長率	日本實質GDP成長率	同時指標綜合指數	
4	大盤股價淨值比	大盤股價淨值比	中國實質GDP成長率	非流動性指標	
5	五大銀行新承做放款利率-購屋貸款	五大銀行新承做放款利率-購屋貸款	日本CPI年增率	落後指標/領先指標	
6	金融業拆款利率	落後指標綜合指數	進口物價總指數	落後指標綜合指數	
7	大盤股價營收比	大盤股價營收比	南韓貨幣市場利率	臺灣物價總指數年增率(未經季調)	
8	日本實質GDP成長率	非流動性指標	台北市房租指數	大盤股價淨值比	

表7 各模型重要變數($h = 12$)

排序	LASSO	EN	Adaptive LASSO	Adaptive EN	目標因子
1	美國聯邦資金市場利率	美國聯邦資金市場利率	美國聯邦資金市場利率	美國聯邦資金市場利率	台灣實質GDP年增率
2	CPI年增率(經季調)	CPI年增率(經季調)	日本CPI年增率	CPI年增率(經季調)	中國CPI年增率
3	日本CPI年增率	日本CPI年增率	CPI年增率(經季調)	美中匯率	MIB年增率
4	美國CPI年增率	美中匯率	中國CPI年增率	日本CPI年增率	CRB期貨指數
5	中國CPI年增率	中國CPI年增率	美國CPI年增率	美國CPI年增率	美國實質GDP成長率
6	美中匯率	美國CPI年增率	美中匯率	中國CPI年增率	台灣十年期公債利率
7	CRB期貨指數	五大銀行新承做放款利率-購屋貸款	CRB期貨指數	五大銀行新承做放款利率-購屋貸款	美國長(十年)短(五年)天期利差
8	五大銀行新承做放款利率-購屋貸款	CRB期貨指數	五大銀行新承做放款利率-購屋貸款	痛苦指數	M2年增率
排序	套袋樹	隨機森林	LASSO-套袋樹	隨機森林-LS	
1	日本CPI年增率	日本CPI年增率	日本CPI年增率	CPI年增率(經季調)	
2	中國實質GDP成長率	五大銀行新承做放款利率-購屋貸款	痛苦指數	日本CPI年增率	
3	五大銀行新承做放款利率-購屋貸款	中國實質GDP成長率	中國實質GDP成長率	五大銀行新承做放款利率-購屋貸款	
4	金融業拆款利率	金融業拆款利率	五大銀行新承做放款利率-購屋貸款	美國CPI年增率	
5	MIB年增率	商業本票31-90天期次級市場利率	M2年增率	躉售物價總指數年增率(未經季調)	
6	商業本票31-90天期次級市場利率	MIB年增率	Delta_CoVar	MIB年增率	
7	CPI年增率(經季調)	落後指標綜合指數	中國CPI年增率	領先指標綜合指數	
8	落後指標綜合指數	CPI年增率(經季調)	CPI年增率(經季調)	落後指標綜合指數	

表8列出向前24個月的預測模型選擇之重要變數，其中LASSO模型家族所選擇的重要變數主要由國際因素相關變數構成，此類模型選擇的國際因素包含日本CPI年增率及美日匯率等，顯示日本經濟對我國通膨預測有相當程度的解釋能力；國內因素則有M2年增率、台灣十年期公債利率、及進、出口物價總指數等。相較於LASSO模型家族，非線性機器學習模型及組合模型選擇的國際因素相關變數較少，包括中國CPI年增率與實質GDP成長率，以及美國CPI年增率。以向前24個月的通膨預測來說，此類模型大多選擇領先、同時、落後指標綜合指數等景氣指標、台灣十年期公債利率及商業本票31-90天期次級市場利率等市場利率，以及非流動性指標。

綜觀所有實證結果，時間序列模型僅在向前一期預測表現尚可。LASSO模型家族、因子模型、非線性機器學習模型及組合模型均在短期及中長期預測表現較佳。其中，非線性機器學習模型之隨機森林模型，在短期及中長期預測有著優良且最為穩定的表現。

藉由觀察不同預測模型在不同預測期間

下所選擇對通膨預測有解釋能力的變數，可歸納出4個重點：(1) 台灣通膨深受國際因素及進出口貿易影響，主要貿易對手國之CPI年增率、實質GDP成長率、匯率及其政策利率等變數，以及進出口物價總指數大多為重要變數，惟就機器學習模型及組合模型而言，此一影響隨預測期間增加而有所下降。(2) 國內因素對通膨預測較有貢獻者，除前一年經季調的CPI年增率外，尚有貨幣總計數相關變數如M1B、M2年增率，市場利率如金融業拆款利率、台灣十年期公債利率及商業本票31-90天期次級市場利率，產業相關如批發業、零售業營業額年增率，供給面變數如躉售物價總指數年增率等。(3) 其他偶經選擇為重要變數之國內因素，如台北市房租指數、五大銀行新承做放款利率-購屋貸款等房市變數，對通膨預測亦有一定解釋能力。(4) 指數類型變數富含通膨預測資訊內涵，如非線性機器學習模型及組合模型在中長期通膨預測時經常選擇領先、同時、落後指標綜合指數等景氣指標，而痛苦指數、非流動性指標等亦經常位列於重要變數，顯示不同指標涵蓋所屬市場的重要訊息。

伍、結 論

本文應用總體經濟及金融市場大數據資料，提出多種機器學習方法預測台灣通膨率，實證結果發現，在向前一期通膨預測方面，LASSO模型家族、非線性機器學習模型及組合模型的預測表現接近，均較基準時間序列模型佳。而在短期或是中長期通膨率預測方面，非線性機器學習模型預測誤差較小，其中以隨機森林的預測表現最佳，明顯改善基準模型預測表現，與Medeiros et al. (2021)的實證結果一致。其他模型如LASSO模型家族、因子模型及組合模型亦有良好的預測表現。

本文所使用的各種機器學習方法中，以非線性機器學習方法之預測表現最佳，可能原因為這些方法有效地捕捉經濟體系中重要總體經濟變數與通膨之間的非線性關係，故能提升通膨預測準確性。本文建議，央行可運用各種機器學習方法作為通膨預測工具，除可改善通膨預測表現外，亦可借助其變數選擇機制，篩選出對通膨預測有重要影響性的關鍵變數，進而加深對通膨傳遞管道的瞭解，據以採取妥適的貨幣政策，以達到央行穩定物價的目標。

表8 各模型重要變數($h = 24$)

排序	LASSO	EN	Adaptive LASSO	Adaptive EN	目標因子
1	台灣十年期公債利率	台灣十年期公債利率	台灣十年期公債利率	台灣十年期公債利率	M1B 年增率
2	美日匯率	台灣實質GDP 年增率	M2 年增率	非流動性指標	M2 年增率
3	非流動性指標	非流動性指標	美日匯率	日本CPI 年增率	台灣實質GDP 年增率
4	日本CPI 年增率	日本CPI 年增率	出口物價總指數	出口物價總指數	美國長(十年)短(五年)天 期利差
5	M2 年增率	出口物價總指數	非流動性指標	台灣實質GDP 年增率	美國CPI 年增率
6	出口物價總指數	南韓貨幣市場利率	日本CPI 年增率	M2 年增率	勞動參與率
7	台灣實質GDP 年增率	進口物價總指數	台灣實質GDP 年增率	美日匯率	美國二年期公債殖利率
8	大盤股價淨值比	M2 年增率	大盤股價淨值比	進口物價總指數	中國CPI 年增率
排序	套袋樹	隨機森林	LASSO-套袋樹	隨機森林-LS	
1	中國實質GDP 成長率	中國實質GDP 成長率	中國CPI 年增率	落後指標綜合指數	
2	中國CPI 年增率	落後指標綜合指數	勞動參與率	躉售物價總指數年增率 (未經季調)	
3	台灣十年期公債利率	領先指標綜合指數	台灣十年期公債利率	非流動性指標	
4	金融帳-資產-其他投資	同時指標綜合指數	信用對產出佔比	五大銀行新承做放款利 率-購屋貸款	
5	台灣全體貨幣機構存放 比	中國CPI 年增率	美國聯邦資金市場利率	落後指標/領先指標	
6	落後指標綜合指數	商業本票31-90天期次級 市場利率	非流動性指標	領先指標綜合指數	
7	非流動性指標	台灣全體貨幣機構存放 比	中國實質GDP 成長率	同時指標綜合指數	
8	五大銀行新承做放款利 率-購屋貸款	金融帳-資產-其他投資	美國CPI 年增率	金融帳-資產-其他投資	

參考文獻

中文文獻

- 朱浩榜(2019)，「探討台灣食物類價格及其與通膨情勢之關係」，中央銀行季刊，第四十一卷第二期，頁5-32。
- 朱浩榜(2023)，「探討臺灣的菲利浦曲線及其通膨預測能力」，臺灣經濟預測與政策，第五十三卷第二期，頁51-103。
- 林依伶、陳佩玕(2021)，「連鎖法衡量下之台灣短中期通膨與經濟成長預測」，中央銀行季刊，第四十三卷第一期，頁7-48。
- 陳佩玕(2013)，「台灣地區短期通貨膨脹率之預測」，中央銀行季刊，第三十五卷第一期，頁63-90。
- 蕭宇翔、繆維正(2021)，「高頻物價數據進行通膨預測」，經濟論文叢刊，第四十九卷第三期，頁371-414。

英文文獻

- Almosova, A., and N. Andresen (2023), "Nonlinear inflation forecasting with recurrent neural networks," *Journal of Forecasting*, 42(2), 240-259.
- Amihud, Y. (2002), "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, 5(1), 31-56.
- Araujo, G. S., and W. P. Gaglianone (2023), "Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models," *Latin American Journal of Central Banking*, 4, 100087.
- Bai, J., and S. Ng (2008), "Forecasting economic time series using targeted predictors," *Journal of Econometrics*, 146, 304-317.
- Barardehi, Y. H., Bernhardt, D., Ruchti, T. G., and M. Weidenmier (2021), "The night and day of Amihud's (2002) liquidity measure," *The Review of Asset Pricing Studies*, 11(2), 269-308.
- Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., and N. Koenigstein (2023), "Forecasting CPI inflation components with hierarchical recurrent neural networks," *International Journal of Forecasting*, 39(3), 1145-1162.
- Baybuza, I. (2018), "Inflation forecasting using machine learning methods," *Russian Journal of Money and Finance*, 77(4), 42-59.
- Breiman, L. (1996), "Bagging predictors," *Mach. Learn.*, 24, 123-140.
- Breiman, L. (2001), "Random forests," *Mach. Learn.*, 45, 5-32.
- Breiman, L., Friedman, J. H., Olshen, R. A., and C. J. Stone (1984), "Classification and regression trees," Belmont, CA: Wadsworth. *International Group*, 432, 151-166.
- Choudhary, M. A., and A. Haider (2012), "Neural network models for inflation forecasting: An appraisal," *Applied Economics*, 44(20), 2631-2635.
- Esquivel Monge, M. (2007), "Performance of artificial neural networks in forecasting Costa Rican inflation," Department of Economic Research, Central Bank of Costa Rica DEC-DIE-029-2009-DI..
- Hoerl, A. E. (1962), "Application of ridge analysis to regression problems," *Chemical Engineering Progress*, 58, 54-59.
- Hoerl, A. E., and R.W. Kennard (1970), "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, 12(1), 55-67.
- Inoue, A., and L. Kilian (2008), "How useful is bagging in forecasting economic time series? a case study of U.S. CPI inflation," *Journal of the American Statistical Association*, 103, 511-522.
- Jorion, P., and F. Mishkin (1991), "A multicountry comparison of term-structure forecasts at long horizons," *Journal of Financial Economics*, 29(1), 59-80.

- Kohlscheen, E. (2021), “What does machine learning say about the drivers of inflation?” *BIS Working Papers* n. 980.
- Medeiros, M., and E. Mendes (2016), “ ℓ_1 -Regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors,” *Journal of Econometrics*, 191, 255–271.
- Medeiros, M., Vasconcelos, G., and E. H. de Freitas (2016), “Forecasting brazilian inflation with high dimensional models,” *Brazilian Review of Econometrics*, 36 (2), 223–254.
- Medeiros, M. C., Vasconcelos, G., Veiga, Á, and E. Zilberman (2021), “Forecasting inflation in a data-rich environment: The benefits of machine learning methods,” *Journal of Business & Economic Statistics*, 39, 98–119.
- Mishkin, F. S. (1989), “The information in the longer maturity term structure about future inflation,” *National Bureau of Economic Research Cambridge, Mass., USA*.
- Rodríguez-Vargas, A. (2020), “Forecasting Costa Rican inflation with machine learning methods,” *Latin American Journal of Central Banking*, 1, 100012.
- Sims, C. A. (1980), “Comparison of interwar and postwar business cycles: Monetarism reconsidered,” *National Bureau of Economic Research Cambridge, Mass., USA*.
- Stock, J., and M. Watson (1999), “Forecasting inflation,” *Journal of Monetary Economics*, 44, 293–335.
- Stock, J., and M. Watson (2002), “Macroeconomic forecasting with diffusion indexes,” *Journal of Business and Economic Statistics*, 20, 147–162.
- Tibshirani, R. (1996), “Shrinkage and selection via the Lasso,” *Royal Statistical Society*, 58(1), 267–288.
- Tobias, A., and M. K. Brunnermeier (2016), “CoVaR,” *The American Economic Review*, 106(7), 1705.
- Zou, H. (2006), “The adaptive lasso and its oracle properties,” *J. Am. Stat. Assoc.*, 101, 1418–1429.
- Zou, H., and T. Hastie (2005), “Regularization and variable selection via the elastic net,” *Royal Statistical Society*, 67(2), 301–320.
- Zou, H., and H. H. Zhang (2009), “On the adaptive elastic-net with a diverging number of parameters,” *Annals of Statistics*, 37(4), 1733–1751.

Forecast Inflation in Taiwan with the Application of Big Data Analysis

Hsin-Yi Lin*, Yu-Min Yen**, Jin-Huei Yeh***

Abstract

This study constructs a comprehensive database comprising 93 macroeconomic and financial variables in Taiwan from January 2000 to December 2023. We utilize a variety of forecasting methodologies, including time series models, the LASSO family of models, factor models, nonlinear machine learning methods, and ensemble models, to predict Taiwan's inflation rate. The empirical results indicate that for one-period-ahead inflation rate forecasts, the performance of linear and nonlinear machine learning and ensemble models is comparable and surpasses that of traditional time series models. In both short-term and medium- to long-term inflation rate forecasts, nonlinear machine learning models exhibit superior performance, likely due to the potential nonlinear relationships between key macroeconomic variables and inflation.

The selection of the critical variables by the forecasting models highlights their significant influence on Taiwan's inflation rate. International factors, particularly those related to import and export trade, play an important role. Variables linked to major trading partner countries consistently provide rich predictive information for inflation. Domestic factors such as monetary aggregates, interest rates, wholesale and retail sales, and supply-side related variables contribute substantially to inflation forecasting. Moreover, housing and stock market indexes are also informative and provide valuable insights into inflation forecasting.

Keywords: Inflation Rate, Machine Learning, Random Forest, Factor Model

JEL classification code: C22, E31

* Department of Economics, National Chengchi University
** Department of International Business, National Chengchi University
*** Department of Finance, National Central University

