

學術論著

利用機器學習和隨機過程預測房價：以台灣房市為例

Predicting Housing Prices via Machine Learning and Stochastic Processes: A Case Study of the Taiwan Property Market

王文楷*

Wen-Kai Wang*

摘要

本文介紹兩個預測房價的方式：長短期記憶模型，以及隨機森林結合隨機微分方程。使用隨機微分方程的主要原因，是為了補抓近期的房市交易人情緒是否過熱或過冷，進而影響近期未來房市，並依此設計相關的指標。長短期記憶模型的預測方式，主要利用具有時間序列性質的資料，直接進行預測；而隨機森林則結合隨機微分方程，以上述指標進行分類。利用台灣房市的實證資料，本文發現兩個預測方式，皆有不錯的表現；而隨機森林結合隨機微分方程的方式，在本文測試方式下優於長短期記憶模型，此代表房市近期情緒對於預測房價可能具有一定效果。

關鍵詞：房價、長短期記憶、隨機森林、情緒、隨機微分方程

ABSTRACT

This study introduces two methods for predicting house prices: Long Short-Term Memory model and Random Forest combined with stochastic differential equations. The main reason for using stochastic differential equations is to capture the sentiments of market participants in the short term and subsequently affect the property market in near future. This enables to determine whether recent property market is overly bullish or bearish, and to design relevant indicators based on this information. The prediction method of the long short-term memory model mainly utilises time-series data for direct predictions, while the random forest combines with stochastic differential equations to perform classification using the aforementioned indicators. Using empirical data from the Taiwan property market, it is found that both prediction methods perform well. However, the random forest combined with stochastic differential equations outperforms the long short-term memory model in the testing conducted in this article, which demonstrates that the long and short-term sentiments of the market may have a certain effect on predicting housing prices.

Key words: Housing Price, Long Short-Term Memory, Random Forest Sentiment, Stochastic Differential Equation

(本文於2024年3月18日收稿，2024年6月20日審查通過，實際出版日期2025年6月)

* 國立高雄大學財務金融學系教授

Professor, Department of Finance, National University of Kaohsiung, Taiwan

E-mail: wkwang@nuk.edu.tw

感謝國科會研究計畫經費補助(NSTC 112-2410-H-390-019 -)

一、前言與文獻探討

房地產長久以來被視為國家經濟的火車頭，房市趨勢上升，除帶動不同產業的發展，進而影響國家經濟外，也使得房產持有者的不動產價值增加。然而，近年來原物料受到國際因素影響，導致原物料成本快速上升，造成薪資成長跟不上萬物齊漲速度；而房價也受到原物料影響，使得民眾購屋壓力增大。部分民眾甚至認為，在少子化的影響下，房價上升是否為合理現象？房價是否被過度炒作？房市是否過熱？由上述說明可見，不管從國家經濟發展、就業市場、民眾財富變動或購屋壓力等感受，都足見房市趨勢變化的重要性；也使得房價預測問題，具有重大研究價值。

有關房價預測相關研究，除房價本身外，也需要考量其它影響房價的因素。影響房價的因素有許多，以利率為例，陳旭昇、陳柏瑜(2024)將台灣做為研究對象，利用1991年至2021年的實證資料，發現以長期來看，台灣的利率政策確實影響房價變化；而作者也提出利率可能藉由其它管道影響房價，例如在低利率下，可能降低房產持有者的使用成本(Himmelberg et al., 2005; Kuttner, 2013)，或是提高承擔風險意願，進而升高購屋機會(Agnello et al., 2018)。另一方面，Vonlanthen(2023)則以瑞士房市為例，發現利率影響了不同類型的房地產類型；而Agnello et al.(2020)也建議可藉由貨幣政策，通過影響利率避免房市過熱發生的可能性。由於有不少文獻發現利率對房價有明顯影響，因此本文即以利率為例，介紹預測房價的方式。

傳統文獻對於房價預測的方式，多是採用計量模型或是時間序列模型，例如hedonic定價模型(Aladwan & Ahamad, 2019; Herath & Maier, 2010; Sirmans et al., 2005; Wing & Chin, 2003; Zakaria & Fatine, 2021)，以及時間序列模型(Chen et al., 2010; Li et al., 2010)。由於使用方法的特性，為了提升預測的正確率，此類模型需要考量較多影響房價的因素。另一方面，一部分文獻以隨機微分方程(stochastic differential equation, SDE)做為房價以及其它因素的模型(陳勤明等，2017；王文楷，2019；王文楷，2023；Chen et al., 2010; Huang et al., 2011; Lee et al., 2012)。Yilmaz & Selcuk-Kestel(2018)指出利用SDE模型化房價与其它因素的主要好處，是可以降低因子數量，只考慮較重要的因子；然而，這類模型主要的應用，並非是預測未來房價，而是定價以房產為標的物的金融商品，例如不動產逆向抵押貸款。

與上述傳統文獻不同，本文所介紹的預測房價方法，是利用近年來在各領域越來越熱門的機器學習模型，本文主要使用隨機森林(random forest, RF)和長短期記憶(long short-term memory, LSTM)。前者由Breiman(2001)提出，其在分類問題上有不錯的效果；而後者則是Hochreiter & Schmidhuber(1997)所發展的一種深度學習模型，其在預測具有時間序列特性的資料上有不錯的表現。機器學習模型在處理資料間的非線性關係上，具有強大的優勢。事實上，機器學習模型在財經領域已有文獻探討相關應用(趙啟方等，2022；Fischer & Krauss, 2018；Hsu, 2021; Krauss et al., 2017; Rather et al., 2015)。Rather et al.(2015)利用遞迴神經網路(recurrent neural networks, RNN)模型進行股價報酬率的預測，並且與兩個線性模型：自我迴歸移動平均模型(autoregressive moving average model)以及指數平滑模型(exponential smoothing model)的結果進行比較，並發現RNN在預測結果上優於所選的兩個線性模型。作者更進一步以適當的權重，將三個模型的預測結果進行組合，發現預測結果更進一步提升。

Krauss et al.(2017)利用了深度神經網路(deep neural networks, DNN)、梯度提升樹(gradient-

boosted trees)、RF，以1992年至2015年之間，標準普爾500指數(S&P 500)的成分股作為測試。藉由特定的投資組合設計，作者發現結合這三個模型所得到的投資組合，有不錯的日報酬率表現。另一方面，Fischer & Krauss(2018)指出LSTM適合用於處理具有時間序列的資料。延續Krauss et al.(2017)的測試，作者也利用1992年至2015年之間S&P 500中的成分股，發現LSTM的表現在未扣除交易成本前，日報酬率和夏普比值(Sharpe ratio)皆優於三個無記憶分類的模型：RF、DNN、羅吉斯迴歸(logistic regression)。然而作者進一步分析後發現，在2010年後LSTM的表現在扣除成本下並不如預期。作者推測可能原因為成分股的高波動與短期反轉效應，因此以此設計對應策略，修正LSTM在2010年後的績效表現。

Hsu(2021)將RF應用於選股問題，作者將基本面分析指標(如：市值和每股盈餘)、技術分析指標(如：RSI和MACD)、動能等，設計為RF的特徵。由於預測股價報酬率相當困難，作者將報酬率分為10組，並改為預測股價報酬率組別。同時因為RF的特性，作者也提出如何處理原始具有時間序列現象的資料，以及重要特徵的篩選方式。作者分別以2014年至2018年間的S&P 500，以及富時100指數(FTSE 100)中的成分股做為實證測試，發現在不同績效指標(如：累積報酬率和夏普比值)下，利用RF選股皆能優於指數的表現。另一方面，趙啟方等(2022)則利用RF，對選定的選擇權策略勝率進行預測。作者發現隨機森林在預測選定的策略勝率上，優於統計方法的預測結果。

而在不動產領域方面，也有文獻利用機器學習方式進行房價或是相關因素的預測(彭蒂菁，2021；蔡繡容、夏政瑋，2023；Annamoradnejad & Annamoradnejad, 2022; Tekouabou1 et al., 2024; Wandhe et al., 2023)。Annamoradnejad & Annamoradnejad(2022)回顧了機器學習對預測房價的相關研究，並提出預測房價應當考量價格的多年波動狀況，以及考量外部經濟變數；而Wandhe et al.(2023)則利用機器學習模型，在考慮房產面積、周邊環境、建築年齡、樓層數、可用房間數和使用面積等因素下，預測未來房價上漲的機率，提供使用者更好的交易機會。另一方面，Tekouabou1 et al.(2024)則發現使用機器學習預測房價的相關研究，有明顯增加的趨勢。作者認為機器學習方法應用於房地產問題需要克服多個重大挑戰，如：優化預測性能、解釋模型以及解決設計模型的複雜性問題。

在台灣的相關研究方面，彭蒂菁(2021)針對未來台灣將邁入的高齡化社會，研究與年長者息息相關的醫療可及性如何影響房價。作者藉由決策樹和RF，發現醫療可及性對大台北房價有顯著影響，而此研究結果對醫療資源的分配有相當的貢獻。蔡繡容、夏政瑋(2023)則是使用LSTM，對房價和交易量進行預測。作者以GARCH模型做為比較對象，並使用2005年第一季至2020年第三季的台北、新北、全國的國泰房價指數資料，以及經濟和非經濟變數進行預測。作者同時測試了三種LSTM的訓練方式：無滾動重新訓練、滾動一次、兩次重新訓練，發現三種方式的預測效果皆優於GARCH模型，另外LSTM則會隨著滾動重新訓練的次數增加而有所提升。

由這些文獻可見機器學習的熱門程度，以及應用的廣度。然而，雖然已有機器學習被用來預測房價的相關文獻，但本文與這些文獻主要不同之處，是結合了SDE的應用。SDE為一種隨機過程模型，而近年已有少量文獻結合隨機過程和機器學習的應用介紹(Chen, 2023; Johansson, 2022)，例如Johansson(2022)介紹如何利用機器學習模型估計SDE的參數。本文利用SDE的主要原因，是如同Yilmaz & Selcuk-Kestel(2018)中所介紹，SDE在進行模型化一個因素

時，不必加入過多因子提升模型精確度，而僅需要考量重要因子。以選擇權定價公式為例，Black & Scholes(1973)使用了幾何布朗運動(geometric Brownian motion, GBM)對標的物進行模型化，而GBM中的因子僅為資產價格的期望成長率和波動度，而其它影響資產價格因子則視為隨機性；該模型雖簡單，卻使選擇權定價問題得到突破性發展。因此本文認為針對過去尚未發生，但未來可能影響房價的部分變數，如政府打房政策、總體經濟變數或人口變化影響等，在SDE中將其視為隨機性處理。

部分文獻探討房市交易人情緒如何影響房市的變化，並進一步編制情緒指數(江明珠、許秉凱，2019；朱芳妮等，2020；Clayton et al., 2009; Chen & Zeng, 2014; Hui et al., 2017)。本文認為近期房市情緒對於近期的未來房價有一定程度影響，如近期房市若過熱，可能使得近期未來的房價上漲機會增加。但要如何捕捉近期房市交易人情緒？由於房價是由房市中的交易人所決定，因此歷史房價應當隱含交易人在交易時的情緒；故當使用過去的歷史資料估計模型參數時，這些參數便包含了過去一段時期的房市交易人情緒。其主要發想來自於長、短天期波動度的比較。一般而言，短天期波動度反映近期市場的交易人情緒，而長天期波動度則較多反映市場的長期趨勢。因此若短天期波動度遠高於長天期波動度時，代表近期市場可能過度恐慌；另一方面，若短天期波動度遠低於長天期波動度時，則可能為近期市場過度樂觀。本文藉由上述概念，利用長、短期歷史資料估計模型參數，結果分別代表市場長期趨勢和市場近期情緒；後續再以蒙地卡羅法，分別模擬近期未來兩種狀況的變化，交叉比對作為市場近期情緒是否有過熱或過冷現象，並利用其結果作為特徵，用以訓練RF進行房價預測。

本文分別利用LSTM和RF結合SDE介紹房價預測方式，而RF結合SDE為本文主要貢獻，LSTM則作為RF結合SDE的比較對象。總結本文貢獻如下：

1. 學術貢獻方面，本文利用隨機過程模型化房價以及與其相關的因素，並以不同時間長度的歷史資料估計模型參數，試圖捕捉市場長期趨勢和市場近期情緒，利用交叉比對判斷近期房市情緒是否過熱或過冷，並以此結合RF。
2. 實務應用方面，本文參考Hsu(2021)利用有限變數設計大量特徵的作法，並在台灣房市的實證中發現，即便預測房價時僅使用少量因素(房價、利率)，仍可藉由結合SDE的方式，設計出一定數量特徵訓練RF進行分類(詳見第四節)，而實證結果顯示有一定績效。

最後，本文的編寫結構如下：首先第二節介紹本文如何使用LSTM，以及RF結合SDE進行未來房價預測；第三節則利用台灣房市和貸款利率的資料進行測試，特別一提的是，LSTM部分並未與SDE結合應用，僅測試利用收集到的原始資料進行訓練，並做為與RF結合SDE預測效果的比較對象。而第四節則提供本文結論，以及未來研究可能延伸的方向。

二、機器學習模型的優缺點

如上節所述，機器學習模型近年來受到不同領域重視，應用範圍也越來越廣；相關財經領域的應用，請見上節的文獻介紹，在此不重述。然而作為一種預測工具，機器學習模型仍有其優缺點，使用上必須注意。由於本文所採用的模型為LSTM和RF，故本節主要介紹這兩種模型的優缺點。

LSTM已被證實在語言翻譯和圖形辨識上有相當好的表現(Graves et al., 2009; Karpathy & Li, 2015; Sutskever et al., 2014; Vinyals et al., 2015)，其改善了傳統的RNN模型在訓練過程中，

容易產生梯度消失或梯度爆炸現象，並且也有效捕捉具有時間序列資料間的相依關係。而LSTM除了在處理具有時間序列資料上有一定的效果外，也可處理具有非線性關係的資料，使其應用性更廣。然而，雖然LSTM改善了梯度消失或梯度爆炸問題，但並非完全解決；訓練過程中的超參數選擇，以及原始資料的預處理方式，使得梯度消失或梯度爆炸仍有可能發生。LSTM的另一個問題，在於模型解釋力不足。由於訓練過程中，資料會不斷經由數個函數和權重矩陣進行運算，導致判斷變數對預測結果影響的難度，目前已有一些文獻探討這方面的議題(Cashman et al., 2018; Lipton, 2016; Strobel et al., 2017)。

在RF方面，該模型由多個決策樹所組成，且訓練每棵決策樹時所使用的資料彼此間互相獨立，因此RF優點之一為避免過度擬合(overfitting) (Breiman, 2001)。而RF在處理分類問題上也有相當好的表現(Fernandez-Delgado et al., 2014; Liaw & Wiener, 2002)，Fernandez-Delgado et al.(2014)比較多個用以處理分類問題的機器學習模型，發現RF優於其它模型。另一方面，使用RF時需要先選擇用以分類的特徵，故相較於LSTM模型解釋力的問題，RF在模型解釋力上較佳。然而，由於RF由多棵決策樹組成，因此一般而言，其在訓練時所需要的時間較長。

雖然LSTM和RF有部分缺點，如模型解釋力、模型訓練時間、過度擬合問題等；但不可否認，兩者在處理非線性資料和預測的準確度上，在不少文獻中被證實有不錯的表現。另一方面，LSTM和RF在財經領域應用中，近年來已有一定文獻探討。不管是預測具有時間序列現象的資料，或是將預測問題轉變為分類問題，皆可利用LSTM和RF處理。由於房價以及與其相關的資料具有時間序列現象，故可嘗試使用LSTM進行預測；而具有時間序列現象的資料，也可在適當處理之下(如轉換為成長率)，去除時間序列現象，並對房價報酬率進行分類。綜合上述說明，本文仍選擇兩模型作為預測方法。

三、預測方法介紹

本節介紹如何利用LSTM和RF進行房價預測，RF的預測方式將結合隨機過程，而LSTM在本文中主要作為RF結合隨機過程的比較對象。假設使用 n 種類型資料進行預測(如房價、利率、人口數等)，並且每隔 $\Delta t > 0$ 收集一組資料，時間點為 $t_i = (i - 1)\Delta t$ 、 $i = 1, 2, \dots, m$ 。令 X 為一個 $m \times n$ 的矩陣，而 $X_{i,j}$ 為第 i 行和第 j 列的元素，其表示時間點 t_i 時，第 j 種資料的數值。以下為LSTM和RF預測方法的介紹。

(一) 方法一：LSTM

LSTM所處理的是時間序列的資料，因此在預測和測試時，不需要打亂 X 的列排序。假設要預測 t_i 時的房價，使用的資料為時間點為 $t_{i'-\tilde{n}+1}$ 、 $t_{i'-\tilde{n}+2}$ 、 \dots 、 $t_{i'}$ ，共 \tilde{n} 組資料。由於訓練LSTM時需要進行最佳化，因此為了加快收斂，以及避免梯度消失或梯度爆炸，需要將 X 中的數值進行正規化(normalised)，使其皆介於-1至1之間。本文參考Singla, Duhan & Saroha(2022)所介紹MinMax正規化方式，並將其延伸如下：若 $\tilde{X}_{i,j}$ 為 $X_{i,j}$ 標準化後的數值，其為：

$$\tilde{X}_{i,j} = \alpha_j \frac{X_{i,j} - X_j^{\min}}{X_j^{\max} - X_j^{\min}} \dots \dots \dots (1)$$

其中 $\alpha_j \in [0,1]$ 、 X_j^{\max} 和 X_j^{\min} 分別為第 j 種資料的最大值和最小值(註1)。特別一提，式(1)可將原始資料轉換至0至1之間。此與蔡繡容、夏政璋(2023)的正規化方式不同；另一方面， α_j 的數值作為一個超參數，可使訓練LSTM時更有彈性，而下一節實證結果中，可發現其對預測結果將有所影響。

有關訓練LSTM時，其所面對數學問題為一個最佳化問題(註2)，主要目標為找出作用於資料上的最適權重矩陣，使模型在預測時，能夠適當的運用過去資料。令 h_i 為 t_i 時的房價，而 \tilde{h}_i 為利用式(1)正規化後的房價；本文訓練LSTM的最佳化問題轉換方式如下：首先將 $t_{i-\tilde{n}+1}$ 、 $t_{i-\tilde{n}+2}$ 、 \dots 、 t_i 時的正規化資料，重新排列為一行。令：

$$Z_{i_1, j}^{i_2, j} = \begin{pmatrix} \tilde{X}_{i_1, j} \\ \tilde{X}_{i_1+1, j} \\ \vdots \\ \tilde{X}_{i_2, j} \end{pmatrix} \dots\dots\dots (2)$$

為由 t_{i_1} 至 t_{i_2} 所排列的第 j 種資料正規化數值，定義：

$$\tilde{Z}_{i-\tilde{n}+1}^i = \begin{pmatrix} Z_{i-\tilde{n}+1, 1}^{i, 1} \\ Z_{i-\tilde{n}+1, 2}^{i, 2} \\ \vdots \\ Z_{i-\tilde{n}+1, n}^{i, n} \\ \tilde{h}_{i-1}^{predict} \\ 1 \end{pmatrix} \dots\dots\dots (3)$$

其中 $\tilde{h}_{i-1}^{predict}$ 為LSTM所預測 t_{i-1} 時的正規化房價， $\tilde{Z}_{i-\tilde{n}+1}^i$ 即為預測 t_i 時使用的所有正規化資料矩陣。LSTM共有三個閥門(gate)，依作用順序分別為：遺忘閥(forget gate)、輸入閥(input gate)、輸出閥(output gate)，其中除輸入閥需要兩個權重矩陣外，其於兩閥門皆由一個權重矩陣決定預測用資料重要性。各個閥門作用依序如下：

1. 遺忘閥利用一個權重矩陣 W^{forget} ，決定當前資料輸入中，有多少比例必須被保留，並與前一期所保留的記憶狀態(memory state)進一步組合。
2. 輸入閥利用兩個權重矩陣 W^{input} 、 \hat{W} ，分別決定當前輸入資料重要性並進行運算後，與上述遺忘閥結果進行組合，成為更新後記憶狀態。更新後記憶狀態，將包含一部份當前輸入資料，以及過去需要被保留的輸入資料。
3. 輸出閥利用一個權重矩陣 W^{output} ，作用於預測用的當前輸入資料，並與更新後的記憶狀態進行組合，成為最新預測結果。因此最新預測結果，將包含當前輸入資料，以及過去需要被保留的輸入資料。

在上述介紹中， W^{forget} 、 W^{input} 、 W^{output} 、 \hat{W} 都是 $1 \times (n\tilde{n} + 2)$ 的列矩陣，並分別作用於式(3)，其作用後結果為一個數值。另一方面，為了避免梯度消失或是梯度爆炸發生，造成無法進行最佳化，在各權重矩陣作用於式(3)後，LSTM進一步使用sigmoid函數和一個雙曲函數進行資料的壓縮。sigmoid函數定義為：

$$S(x) = \frac{1}{1+e^{-x}} \dots \dots \dots (4)$$

而雙曲函數定義為：

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \dots \dots \dots (5)$$

根據上述LSTM運算步驟的介紹，在預測 t_i 的房價時，需要輸入資料 $\tilde{Z}_{i'-\tilde{n}+1}^{i'}$ ，以及 $t_{i'-1}$ 時的記憶狀態 $m_{i'-1}$ 。若輸入資料 $\tilde{Z}_{i'-\tilde{n}+1}^{i'}$ 經由各權重矩陣運算後為：

$$G^k(\tilde{Z}_{i'-\tilde{n}+1}^{i'}) = W^k \tilde{Z}_{i'-\tilde{n}+1}^{i'} \dots \dots \dots (6)$$

以及

$$\hat{G}(\tilde{Z}_{i'-\tilde{n}+1}^{i'}) = \hat{W} \tilde{Z}_{i'-\tilde{n}+1}^{i'} \dots \dots \dots (7)$$

其中 $k = forget, input, output$ ，則 t_i 時的記憶狀態 $m_{i'}$ ，以及預測正規化房價 $\tilde{h}_{i'}^{predict}$ 分別為：

$$m_{i'} = m_{i'-1} S(G^{forget}(\tilde{Z}_{i'-\tilde{n}+1}^{i'})) + \tanh(\hat{G}(\tilde{Z}_{i'-\tilde{n}+1}^{i'})) S(G^{input}(\tilde{Z}_{i'-\tilde{n}+1}^{i'})) \dots \dots \dots (8)$$

和

$$\tilde{h}_{i'}^{predict} = \tanh(m_{i'}) S(G^{output}(\tilde{Z}_{i'-\tilde{n}+1}^{i'})) \dots \dots \dots (9)$$

在式(8)中， $m_{i'-1} S(G^{forget}(\tilde{Z}_{i'-\tilde{n}+1}^{i'}))$ 為遺忘閥中 $t_{i'-1}$ 時的記憶狀態 $m_{i'-1}$ ，根據當前輸入資料 $\tilde{Z}_{i'-\tilde{n}+1}^{i'}$ 而被保留的部分；而另一項 $\tanh(\hat{G}(\tilde{Z}_{i'-\tilde{n}+1}^{i'})) S(G^{input}(\tilde{Z}_{i'-\tilde{n}+1}^{i'}))$ ，則為當前輸入資料 $\tilde{Z}_{i'-\tilde{n}+1}^{i'}$ 應加入新記憶狀態的部分。另一方面，式(9)呈現當前輸入資料 $\tilde{Z}_{i'-\tilde{n}+1}^{i'}$ 以及新記憶狀態，如何決定最新的預測結果。

最後，為了求出四個最適權重矩陣，利用 $t_{\tilde{n}}$ 、 $t_{\tilde{n}+1}$ 、 \dots 、 t_i 時，LSTM的預測正規化房價和真實正規化房價，定義下列損失函數(loss function)：

$$\sum_{i=\tilde{n}}^{i'} (\tilde{h}_i^{predict} - \tilde{h}_i)^2 \dots \dots \dots (10)$$

訓練目標則是找出最小化損失函數的四個權重矩陣。此問題需要使用數值最佳化方法，例如最深梯度下降(deepest gradient method)或是隨機梯度下降(stochastic gradient method)等；另外，為了預測 $t_{\tilde{n}}$ 時的房價，需要利用到 $t_{\tilde{n}-1}$ 時記憶狀態和預測正規化房價，本文將兩者皆假設為0，即： $m_{\tilde{n}-1} = \tilde{h}_{\tilde{n}-1}^{predict} = 0$ 。

(二) 方法二：RF結合隨機過程

由於LSTM是一種深度記憶模型，其訓練過程中需要利用sigmoid函數和雙曲函數進行資料壓縮，因此難以分析各種輸入資料的重要性，以及預測當期時，過去資料的重要性。如第二節所述，深度學習模型的缺點之一，即為模型缺乏解釋力，較難追蹤各個自變數對應變數

的重要程度；另一方面，RF則需要使用者先選擇適合的特徵，並且容易檢查使用特徵的重要性，因此RF在結果解釋上較為容易。然而與LSTM不同，由於RF訓練過程中使用到拔靴法(bootstrap method)，RF處理的資料不應具有時間序列特性，因此在使用資料矩陣 X 時，需要先進行處理消除時間序列現象；而RF處理的問題主要為分類問題，因此需要設計適當方式，將沒有時間序列現象的分類問題，轉化為具有時間序列現象的預測問題。RF的概念，主要是利用多棵決策樹進行分類，而分類方式，則是由決策樹在每一個節點，以吉尼(Gini)不純度或熵(entropy)所計算的資訊增益量決定使用的特徵，並用其對資料進行分類，並重複數次直至滿足停止分類條件；最後再將每一棵決策樹的分類結果，以投票結果作為RF分類結果。

為了達成上述目的，本文先利用隨機過程，模型化各種資料。我們假設各種資料皆為特定SDE的解，其中第 j 種資料滿足下列SDE：

$$dx_j(t) = f_j(t, x(t))dt + g_j(t, x(t))dW_j(t) \dots\dots\dots(11)$$

其中 $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$ ，而 $W_j(t)$ 為一個Wiener過程。本文利用 X 矩陣中的資料，結合適當方法，如最大似法、最小平方法、廣義動差法等，對式(11)進行模型參數估計。由於估計模型參數使用了歷史資料，本文認為歷史資料中隱含了長期市場趨勢或近期市場交易人的情緒，因此利用不同時期的資料，分別估計模型參數後進行比對，則可視為市場是否在短期內，有情緒過熱或過冷的現象發生，而近期房價將受到近期市場情緒影響。而近期未來房價則可藉由蒙地卡羅法進行模擬，並觀察市場長期趨勢和近期情緒下的房市變化。

假設估計模型參數所使用的資料長度分別為 n_1 和 n_2 ，其中 $n_1 > n_2$ ，而模擬未來房價的時間長度為 n_3 。如上所述，若當前時間點為 t_i ，我們分別利用 t_{i-n_1+1} 至 t_i ，以及 t_{i-n_2+1} 至 t_i 的資料估計模型參數，其各自反應了市場長期趨勢和近期交易人情緒；接下來在市場長期趨勢和近期情緒下，分別模擬 t_{i+1} 至 t_{i+n_3} 的各類資料。為了消除資料時間序列現象，本文對第 j 種資料的模擬結果，進行下面的比較：

$$y_{k,j} = \ln \left(\frac{x_{k,j}^{long}}{x_{k,j}^{short}} \right) \dots\dots\dots(12)$$

其中 $x_{k,j}^{long}$ 和 $x_{k,j}^{short}$ 分別為利用 n_1 和 n_2 期估計模型參數，所模擬第 j 種資料於為 t_k 時的數值。 $y_{k,j}$ 為市場長期趨勢和近期情緒的比較，其值越接近0時，代表長期趨勢和近期情緒越接近；另一方面，其值越小代表近期市場情緒越熱，反之代表近期市場情緒越冷，同時這樣的設計也可消除資料時間序列的特性。由於 $y_{k,j}$ 為隨機，本文利用蒙地卡羅法模擬多次後，對 $y_{k,j}$ 分別估計最小值、平均值、最大值，做為分類用特徵，其代表市場長期趨勢和近期情緒於未來三個月內，分別在上方風險、平均狀況、下方風險三種情況的比較。上述說明歸納如下：

1. 發生上方風險時，若近期情緒下所預測的近期未來房價，足夠高於長期趨勢下所預測的近期未來房價，則為近期市場過熱。
2. 發生下方風險時，若近期情緒下所預測的近期未來房價，足夠低於長期趨勢下所預測的近期未來房價，則為近期市場過冷。
3. 若近期情緒下所預測的近期未來房價，與長期趨勢下所預測的近期未來房價有較大差距，則平均上而言，近期房市有過冷或過熱現象。

因此本文藉由估算式(12)，訓練RF利用估算結果進行分類。另一方面，此作法也將影響未來房價的可能因素，如政府房市政策、總經變數、少子化的影響等，皆以隨機性表現並納入考量中。根據上述說明，特徵的個數共有 $3n$ 種。

上述為建立特徵的方式，有關利用RF進行房價預測介紹如下：首先在 t_i 時，計算房價於 t_i 至 t_{i+n_3} 時的平均成長率，即：

$$\bar{h}_i = \left(\prod_{k=i}^{n_3-1} \frac{h_{k+1}}{h_k} \right)^{\frac{1}{n_3}} - 1 \dots\dots\dots (13)$$

使用平均成長率而非模擬房價的主要原因，是為了消除房價的時間序列性質；而直接利用 t_i 以後的模擬結果，也與預測未來房價目的一致。有關分類方式，選擇兩個平均成長率 r_{min} 和 r_{max} ，滿足 $r_{min} < r_{max}$ ，使得式(13)結果皆介於 r_{min} 和 r_{max} 之間；並將 $[r_{min}, r_{max}]$ 等分為 N 段，定義 G_i 為式(13)界於 $r_{min} + (i - 1)\Delta r$ 和 $r_{min} + i\Delta r$ 之間的集合，其中 $\Delta r = \frac{r_{max} - r_{min}}{N}$ 。為了訓練RF，我們針對所有特徵各設定幾個門檻值，並從中隨機選擇數個，建構二元決策樹分類並重複多次。而訓練完成後，為了預測 t_i 時的房價，若RF利用 t_i 時的特徵進行分類預測，且RF預測結果為某一個 G_k 時，本文利用 G_k 解釋為平均成長率，將房價預測結果設計為：

$$h_{i'-1} \frac{r_{min} + (k-1)\Delta r + r_{min} + k\Delta r}{2} = h_{i'-1} \left[r_{min} + \left(k - \frac{1}{2} \right) \Delta r \right] \dots\dots\dots (14)$$

由於利用RF處理分類問題，誤差定義通常以分類是否成功做為標準，但此誤差設計方式對本文的預測房價問題並不適當。例如房價實際平均成長率組別若為 G_i ，而RF分類結果越接近 G_i 時，應視為預測越精準。因此在進入下一節的實證測試前，本文提供較適當的誤差定義方式，以衡量RF預測的效果，並且提供利用RF預測的建議步驟。首先定義：

$$g_{min} = \min\{i | G_i \neq \Phi, i = 1, 2, \dots, N\} \dots\dots\dots (15)$$

和

$$g_{max} = \max\{i | G_i \neq \Phi, i = 1, 2, \dots, N\} \dots\dots\dots (16)$$

即分類中非空集合的最小和最大編號。則修正後預測誤差為：

$$\frac{| \text{所有實際編號平均} - \text{所有RF分類編號平均} |}{g_{max} - g_{min}} \dots\dots\dots (17)$$

式(17)的設計方式，在僅有兩類結果時，與分類問題的誤差一致。另外在實際測試後，可能因為分類過細，即 N 值過大時，造成預測誤差較小的現象，可能難以判讀預測效果。因此在使用上述的RF預測方式，本文建議步驟如下：

1. 先以僅兩類方式，即 $N = 2$ ，確認RF在分類結果有一定效果，即僅分兩類時的誤差需足夠小。
2. 在上述步驟確認後，再選擇較大的 N 值，重新訓練RF進行預測。

四、實證結果

本節介紹上一節所介紹兩種預測方式的實證結果，本文目標雖為房價預測，但房價資料取得困難，且因地區不同而有所差異，故以房價指數作為房價代理變數；另一方面，雖然上節所介紹RF分類結果為房價平均成長率，但根據式(14)，仍可將其結果轉換為房價預測。另外一提的是，本文所使用程式皆為自行編寫，使用軟體為MatLab。

本文實證測試以台灣房市為例，並以歷史房價和利率，對未來房價進行預測。根據上節中介紹RF需要結合SDE進行預測，本文利用Lee et al.(2012)中的房價和利率模型，模型化歷史房價和利率。若 $x_1(t)$ 和 $x_2(t)$ 分別為時的房價和利率，則兩者分別滿足：

$$dx_1(t) = \mu x_1(t)dt + \sigma x_1(t)dW_1(t) \dots \dots \dots (18)$$

和

$$dx_2(t) = \kappa(\theta - x_2(t))dt + \eta\sqrt{x_2(t)}dW_2(t) \dots \dots \dots (19)$$

有關估計式(18)和式(19)參數，首先式(18)為GBM，利用歷史房價以最大似法估計 μ 和 σ ；而式(19)為Cox-Ross-Ingersoll (CIR)過程，利用歷史利率以最小平方方法估計 κ 、 θ 、 η ；最後再利用Wiener過程定義中的獨立條件，估計 $W_1(t)$ 和 $W_2(t)$ 的相關係數(註3)。

有關實證資料來源，本文使用的房價代理變數為清華安富房價指數，利率則使用中央銀行所發布的五大銀行貸款利率。選擇清華安富指數的主要原因，為該指數提供月資料，而非其它房價指數多以季資料為主。這使得在實證測試上，不會造成資料過少的影響。本文收集2012年8月至2023年8月的房價指數和貸款利率資料，共計133組；而

1. 前109組(2012年8月至2021年8月)做為訓練RF的資料集，包含訓練集和驗證集。
2. 最後24個月(2021年9月至2023年8月)的房價指數則做為測試集。

在訓練模型方面，LSTM採用方式，為每月重新訓練模型後預測下一個月房價，即LSTM的訓練集合會隨時間更新，因此重新訓練頻率較高；而RF僅利用前109組資料進行訓練，因此訓練頻率較低，主要原因是為了避免過度擬合，以及提高預測正確率，需要生產足夠多的決策樹，導致RF訓練過程時間較久。

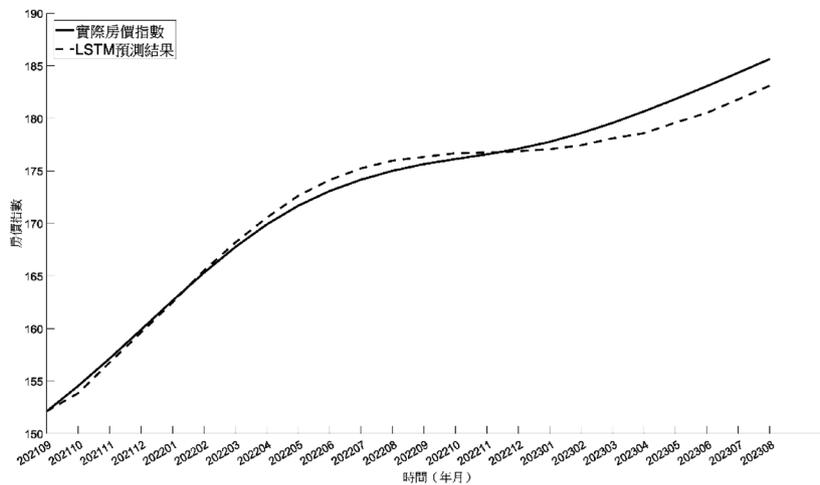
首先呈現LSTM的預測結果，由於本文正規化公式，即式(1)中 α_j 參數可自行調整，為簡化分析，令 $\alpha_1 = \alpha_2 = \alpha$ ；另一方面，用以預測的過去資料長度 \tilde{n} ，本文分別測試四個時間長度(月)的預測能力，並呈現不同 α 下，使得測試期間24個月的相對誤差絕對值之平均最低的結果。表一結果顯示，預測未來一個月房價指數的資料最佳時間長度，為過去9個月的房價指數和貸款利率，而正規化式(1)中的參數 α 則是0.9。本文進一步在圖一中，呈現測試期間(2021年9月至2023年8月)的實際房價指數和LSTM預測房價指數。由圖一中可發現，LSTM的預測結果有不錯的表現。

有關利用RF和SDE預測房價的實證結果，如前所述，本文選擇兩個時間長度，分別估計式(18)和式(19)的模型參數，其反映市場長期趨勢和市場近期情緒。本文測試如下：首先選定RF一組參數組合，並將2012年8月至2021年8月的資料隨機重新排列，以7:3分為訓練集和驗證集。每次由訓練集以拔靴法，隨機選取70%的資料訓練一棵決策數，共生成10,000棵後，再

表一 不同正規化公式參數 α 下，以過去時間長度預測下一個月房價指數的最佳長度、對應的相對誤差範圍、相對誤差絕對值平均

	最佳時間長度(月)	最小相對誤差	最大相對誤差	相對誤差絕對值平均
0.6	15	-0.020	0.025	0.012
0.7	12	-0.013	0.019	0.009
0.8	9	-0.010	0.013	0.007
0.9	9	-0.015	0.007	0.006
1	9	-0.029	-0.003	0.010

註：實證結果由本文整理



圖一 $\alpha = 0.9$ 以及預測時間長度9個月下，2021年9月至2023年8月的實際房價指數和 LSTM預測房價指數

註：實際房價指數來源為清華安富指數

以驗證集檢查該組參數的預測效果，而同一組參數重複上述步驟100次。最後發現不同的時間長度組合，以 $n_1 = 30$ (月)和 $n_2 = 10$ (月)的組合，在預測未來3個月的平均房價指數成長率是否為正，平均上有較佳的表現，預測正確率平均為93.35%，其也符合上節中RF使用建議的第一步驟(註4)。根據RF所訓練的結果，估計市場長期趨勢和近期情緒的較佳時間長度，分別是過去30個月和10個月的實證資料。特別一提的是，上述的測試方式，與一般RF使用的k-folds交叉驗證方式(k-fold cross validation)不同。k-folds法是將資料等分為k份後，每一份輪流做為驗證集，以找出RF的最佳參數組合；而本文由於先將資料的時間序列現象消除，因此可隨機重排訓練集資料，使得所找出的RF參數組合更有可信度。值得一提的是，此處反應短期房市情緒的時間長度，與LSTM預測未來一個月房價所使用的歷史資料長度，兩者相當接近，推測可能原因為LSTM也利用近期市場情緒進行預測。

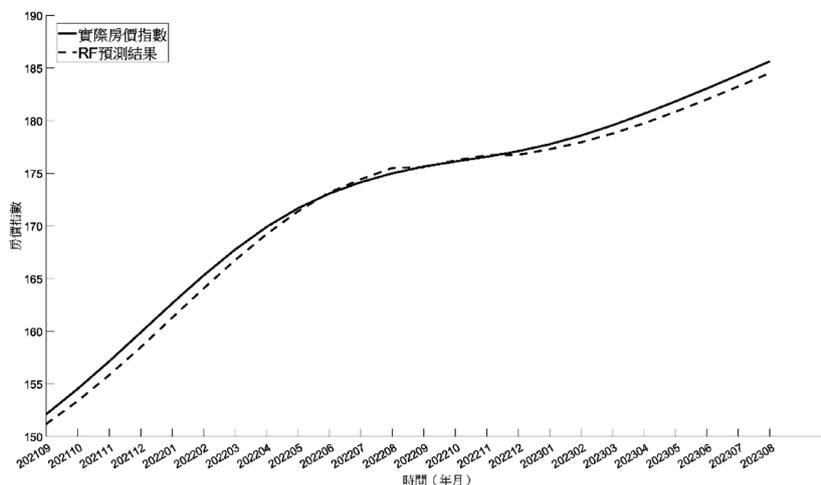
確認RF結合SDE預測未來3個月平均房價成長率是否為正，有足夠高的預測正確率後，本文進一步將分組數提高，並對測試集(2021年9月至2023年8月)的房價指數進行預測。令

$r_{min} = -2\%$ 、 $r_{max} = 2\%$ ，並將 $[-2\%, 2\%]$ 等分為40段。利用2012年8月至2021年8月的資料訓練RF，從中以拔靴法隨機選取70%的資料，訓練每一棵決策樹，共訓練10,000棵決策樹，預測相關結果呈現於表二。表二中顯示分組正確率為0，即實際分組編號和RF預測分組編號皆不一致，另一方面修正後誤差式(17)也高達0.201；然而這是分組較細時，可能出現的現象。由於每組未來3個月平均房價成長率組距為 $\frac{0.02 - (-0.02)}{40} = 0.001$ ，根據修正後誤差，RF預測的未來3個月平均房價成長率，平均上與實際結果差距約為0.4%。若是進一步檢查預測相對誤差，發現其表現較優於LSTM的預測結果。最後，圖二呈現2021年9月至2023年8月的實際房價指數和RF預測房價指數，由圖可見RF結合SDE預測有一定的效果。

表二 利用RF和SDE預測2021年9月至2023年8月的房價指數結果

估計長期情緒預測月數 (n_1)	估計短期情緒預測月數 (n_2)	預測未來房價 平均成長率月數(n_3)	$[r_{min}, r_{max}]$
30	10	3	
分組數	分組正確率	分組編號差絕對值平均	修正後誤差(式(17))
40	0.00%	4.227	0.201
決策樹棵數	最小相對誤差	最大相對誤差	相對誤差絕對值平均
10,000	-0.009	0.004	0.004

註：實證結果由本文整理



圖二 2021年9月至2023年8月的實際房價指數和RF預測房價指數

註：實際房價指數來源為清華安富指數

上述實證測試主要發現RF結合SDE的房價預測，優於本文LSTM僅使用房價和利率為變數的預測效果。為了進一步驗證RF結合SDE的預測效果，本文選擇同樣以台灣房市為實證對象的蔡繡容、夏政瑋(2023)為比較對象。由於蔡繡容、夏政瑋(2023)中所使用的資料為國泰房價指數的季資料，並且測試期間(2018年第4季至2020年第3季)也不同，因此本文僅將預測期間

調整至與蔡繡容、夏政瑋(2023)一致，而房價指數資料仍使用清華安富房價指數。另外，蔡繡容、夏政瑋(2023)中表三的均方根誤差(root-mean-square error, RMSE)結果，判斷為正規化資料之比較，因此需先將其調整為原始資料的RMSE，以利後續比較。利用蔡繡容、夏政瑋(2023)中的式(1)，可得到原始資料RMSE，應為正規化資料RMSE乘上2005年第1季至2018年第3季之間的最大房價指數差(44.7)。

表三 本文模型與蔡繡容、夏政瑋(2023)台灣地區各模型房價預測之RMSE，LSTM^{TS}表示蔡繡容、夏政瑋(2023)所用LSTM模型，而LSTM^W為本文所用LSTM模型

GARCH	LSTM ^{TS}	LSTM ^{TS} -1 st Rolling Window	LSTM ^{TS} -2 nd Rolling Window	LSTM ^W	RF
8.2579	5.2746	0.3576	0.1788	0.7717	0.3449

註：實證資料由本文整理，比對資料來源為蔡繡容、夏政瑋(2023)

表三呈現本文模型(LSTM^W、RF)對台灣地區房價指數預測之RMSE，以及蔡繡容、夏政瑋(2023)所使用模型的預測結果RMSE。由表三可發現，本文所用兩模型在預測績效上皆優於GARCH；並且LSTM^W雖然僅使用房價指數和利率兩變數進行預測，但其預測績效仍優於LSTM^{TS}使用多個變數的效果，而皆不如LSTM^{TS}結合滾動訓練的預測績效。另一方面，RF模型的預測績效接近LSTM^{TS}一次滾動訓練的效果，而不如LSTM^{TS}兩次滾動訓練。但仍需特別一提的是，本文所使用的變數僅有兩個，並且在模型化變數的SDE選擇上，尚有其它可能性，因此仍有修正空間。

最後，由於本文實證資料期間(2012年8月至2023年8月)的房價指數多為上漲趨勢，未來3個月房價平均成長率為正者，比例約為89%；實證資料期間未來3個月房價平均成長率為負者僅有兩個時段：2015年5月至2016年2月，以及2016年9月至2016年11月。為了驗證是否因實證資料結果一致性過高，進而影響模型預測，因此本文進一步測試房市下跌狀況的預測績效。首先，若以2012年8月至2015年4月資料訓練兩個模型，用以預測2015年5月至2016年2月的房價指數，本文LSTM和RF模型的RMSE分別為1.2882和2.0969，可以發現RF模型的預測效果遠不如LSTM，推測原因是RF的預測結果受到實證資料一致性過高影響；另一方面，若以2012年8月至2016年1月的資料訓練模型，用以預測2016年2月至2016年11月(包含第二段房市下跌)的房價指數，本文LSTM和RF兩模型的RMSE分別為0.7703和0.3315，預測表現上有所改善，推測原因是後者的訓練資料中，包含2015年5月至2016年2月房市下跌狀況。因此為了提升預測效果，應當讓訓練資料包含各種不同狀況。

五、結論

近年來AI科技蓬勃發展，各領域陸續嘗試利用相關技術，進行問題探討和資料分析，其中也包括財經領域。而房地產對國家經濟和社會公平等方面有重大影響，其變動不僅帶動相關產業，也關係到人民居住的權益，值得政府重視；因此房市相關因素的預測，一直以來皆是重要議題。本文以台灣房市為例，介紹如何利用深度學習模型LSTM，以及機器學習模型RF結合SDE進行房價指數預測。

在LSTM方面，本文的實證結果發現，其預測表現有一定效果，這點與蔡繡容、夏政璋(2023)的結果相似；主要不同之處，在於本文僅使用兩種因素(房價指數、利率)進行預測，仍達到一定的效果。另一方面，本文使用長、短期資料估計SDE模型參數，並隨後模擬未來變化，試圖捕捉近期房市情緒是否有過熱或過冷現象，並對近期未來房市的影響。本文藉由比對房市長期趨勢和近期情緒下的近期未來房價模擬，判斷近期情緒是否過熱或過冷；而上述發想，則進一步作為特徵，利用RF進行分類和預測。實證結果也發現RF結合SDE在預測上有不錯的效果。總結本文貢獻：

1. 機器學習結合隨機過程的應用相關文獻，目前仍屬少見。本文以不同長度的歷史資料估計隨機過程參數，並模擬近期未來的房市變化，作為判斷房市情緒是否過熱或過冷，並進一步將其設計為RF訓練所需特徵。目前此作法較為罕見，也是本文的重要貢獻之一。
2. 本文所設計實驗方式以台灣房市為例，在實證結果上顯示有一定效果，佐證此預測方式具有實務應用可能性。

特別一提的是，針對部分民眾所質疑房市是否過熱、少子化現象下房價上升是否為合理現象等質疑，本文不管是LSTM或是RF，實證結果發現在2021年9月至2023年8月期間，預測房價和實際房價並未有太大差異，本文認為其結果皆未支持這些質疑。推測可能原因如下：少子化的影響尚未發酵，使得房市長期趨勢與房市近期情緒並未有過大差異，因此對近期未來房價無太大影響。

本文建議未來研究可嘗試將本文所介紹概念應用到其他房市，進行更進一步的驗證。另一方面，本文僅採用兩個因素：房價指數和貸款利率，未來研究也可嘗試加入更多因素，或是採用不同的隨機過程模型化各因素。除了預測房價的應用外，本文提出的研究方法，也可應用於建構房市情緒。在預測正確率足夠高時，若是預測結果與實際結果於一段期間內產生較大差異，推測可能原因是該段期間內，房市有過熱或過冷現象，此時政府機關應審慎檢視房市狀況。最後，上述利用LSTM、RF結合SDE預測房市的方法，以及建構房市情緒的可能應用，皆可提供給政府機關，做為制定相關政策的參考方向。

註 釋

註1： $\alpha_j = 1$ 即為Singla et al.(2022)所介紹MinMax正規化公式。

註2：由於LSTM概念的介紹和圖解有相當多的來源，本文在此不再重述。

註3：上述估計細節，因篇幅關係在此省略。

註4：由於測試的長短時間長度組合數較多，在此省略不同組合的測試結果。

參考文獻

中文部分

王文楷

2019 〈不動產逆向抵押貸款評價與提前解約之分析〉《住宅學報》28(2)：55-73。

2023 〈考量房價與利率關係的不動產逆向抵押貸款定價：利用最大熵原則〉《住宅學報》32(1)：55-73。

Wang, W. K.

2019 “Pricing Reverse Mortgage under the Relationship between Housing Price and Interest Rate: Using the Principle of Maximum Entropy,” *Journal of Housing Studies*. 32(1): 55-73.

2023 “Analyzing the Pricing and Termination of Reverse Mortgages,” *Journal of Housing Studies*. 28(2): 63-80.

江明珠、許秉凱

2019 〈媒體新聞能否預測住房市場？〉《住宅學報》28(2)：37-61。

Chiang, M. C. & B. K. Shiu

2019 “Can the News Media Predict the Housing Market?,” *Journal of Housing Studies*. 28(2): 37-61.

朱芳妮、楊茜文、蘇子涵、陳明吉

2020 〈情緒會影響房市嗎？指數編制與驗證〉《住宅學報》29(2)：35-68。

Chu, F. N., C. W. Yang, T. H. Su & M. C. Chen

2020 “Will Sentiment Influence the Housing Market? Index Compilation and Verification,” *Journal of Housing Studies*. 29(2): 35-68.

陳勤明、洪志興、李宜熹

2017 〈我國不動產逆向抵押貸款評價〉《住宅學報》26(1)：85-100。

Chen, C. M., C. H. Hung & Lee, Y. H.

2017 “The Valuation of Reverse Mortgages in Taiwan,” *Journal of Housing Studies*. 26(1): 85-100.

陳旭昇、陳柏瑜

2024 〈台灣的利率政策與房價〉《經濟論文叢刊》52(4)：377-409。

Chen, S. S. and B. Y. Chen

2024 “Interest Rate Policy and House Prices: Evidence from Taiwan,” *Taiwan Economic Review*. 52(4): 377-409.

彭蒂菁

2021 〈醫療可及性是否左右房價？機器學習之迴歸樹及隨機森林模型的應用〉《應用經濟論壇》109：115-167。

Peng, T. C.

- 2021 “Does Healthcare Accessibility Matter to Housing Prices? Application of Machine-Learning Decision Tree and Random Forest,” *Taiwan Journal of Applied Economics*. 109: 115-167.

趙啟方、王昱嬪、吳牧恩

- 2022 〈運用隨機森林演算法於選擇權量化交易策略〉《中國統計學報》60：69-94。

Chao, C. F., Y. C. Wang & M. E. Wu

- 2022 “Option Quantitative Trading Strategy Based on Random Forest Algorithm,” *Journal of the Chinese Statistical Association*. 60: 69-94.

蔡繡容、夏政瑋

- 2023 〈預測台灣房地產市場趨勢之模型—應用深度學習技術〉《住宅學報》32(2)：21-55。

Tsai, H. J. & C. W. Hsia

- 2023 “Prediction Model for the Trend of Taiwan’s Real Estate Market - Applying Deep Learning Technology,” *Journal of Housing Studies*. 26(1): 85-100.

英文部分

Agnello, L., V. Castro & R. M. Sousa

- 2018 “Economic Activity, Credit Market Conditions, and the Housing Market,” *Macroeconomic Dynamics*. 22(7): 1769–1789.

- 2020 “The Housing Cycle: What Role for Mortgage Market Development and Housing Finance?,” *Journal of Real Estate Finance & Economics*. 61: 607-670.

Aladwan, Z. & M. S. Ahamad

- 2019 “Hedonic Pricing Model for Real Property Valuation via GIS – A Review,” *Civil and Engineering Reports*. 29(3): 34-47.

Annamoradnejad, R. & I. Annamoradnejad

- 2022 “Machine Learning for Housing Price Prediction,” in *Encyclopedia of Data Science and Machine Learning*. 2728-2739. ed. J. Wang, PA: IGI Global.

Black, F. & M. Scholes

- 1973 “The Pricing of Options and Corporate Liability,” *Journal of Political Economy*. 81(3): 637-654.

Breiman, L.

- 2001 “Random Forests,” *Machine Learning*. 45(1): 5-32.

Cashman, D., G. Patterson, A. Mosca, N. Watts, S. Robinson & R. Chang

- 2018 “RNNbow: Visualizing Learning Via Backpropagation Gradients in RNNs,” *IEEE Computer Graphics and Applications*. 38(6): 39-50.

Chen, H., S. H. Cox & S. S. Wang

- 2010 “Is the Home Equity Conversion Mortgage in the United States Sustainable? Evidence from Pricing Mortgage Insurance Premiums and Non-Recourse Provisions Using the Conditional Esscher Transform,” *Insurance: Mathematics and Economics*. 46(2): 371-384.

Chen, N.

- 2023 “Combining Stochastic Models with Machine Learning. In: Stochastic Methods for Modeling and Predicting Complex Dynamical Systems,” *Synthesis Lectures on Mathematics & Statistics*. Springer, Cham. https://doi.org/10.1007/978-3-031-22249-8_10

Chen, P. F. & J. H. Zeng

- 2014 “The Impact of Noise Trader Sentiment on the Housing Market,” *Journal of Management*. 31(3): 245-262.

Clayton, J., D. C. Ling & A. Naranjo

- 2009 “Commercial Real Estate Valuation: Fundamentals Versus Investor Sentiment,” *Journal of Real Estate Finance and Economics*. 38: 5-37.

Fernandez-Delgado, M., E. Cernadas, S. Barro & D. Amorim

- 2014 “Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?,” *Journal of Machine Learning Research*. 15: 3133-3181.

Fischer, T. & C. Krauss

- 2018 “Deep Learning with Long Short-Term Memory Networks for Financial Market Prediction,” *European Journal of Operation Research*. 270(2): 654-669.

Graves, A., M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke & J. Schmidhuber

- 2009 “A Novel Connectionist System for Improved Unconstrained Handwriting Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 31(5): 855-868.

Herath, S. K. & G. Maier

- 2010 “The Hedonic Price Method in Real Estate and Housing Market Research - A Review of the Literature,” *Institute for Regional Development and Environment*. 1-21.

Himmelberg, C., C. Mayer & T. Sinai

- 2005 “Assessing High House Prices: Bubbles, Fundamentals and Misperceptions,” *Journal of Economic Perspectives*. 19(4): 67-92.

Hochreiter, S. & J. Schmidhuber

- 1997 “Long Short-Term Memory,” *Neural Computation*. 9(8): 1735-1780.

Hsu, T. Y.

- 2021 “Machine Learning Applied to Stock Index Performance Enhancement,” *Journal of Banking and Financial Technology*. 5: 21-33.

- Huang, H. C., C. W. Wang & Y. C. Miao
 2011 “Securitization of Crossover Risk in Reverse Mortgages,” *The Geneva Papers on Risk and Insurance-Issues and Practice*. 36(4): 622-647.
- Hui, E. C. M., Z. Dong, S. Jia & C. H. L. Lam
 2017 “How Does Sentiment Affect Returns of Urban Housing?,” *Habitat International*. 64: 71-84.
- Johansson, O.
 2022 “Stochastic Modeling using Machine Learning and Stochastic Differential Equations,” *Master’s Thesis in Engineering Mathematics and Computational Sciences*. Chalmers University of Technology.
- Karpathy, A. & F. F. Li
 2015 “Deep Visual-Semantic Alignments for Generating Image Descriptions,” *arXiv Preprint*. arXiv:1412.2306v2.
- Krauss, C., X. A. Do & N. Huck
 2017 “Deep Neural Networks, Gradient-Boosted Trees, Random Forests: Statistical Arbitrage on the S&P 500,” *European Journal of Operation Research*. 259(2): 689-702.
- Kuttner, K. N.
 2013 “Low Interest Rates and Housing Bubbles: Still No Smoking Gun,” *The Role of Central Banks in Financial Stability*. 159-185.
- Lee, Y. T., C. W. Wang & H. C. Huang
 2012 “On the Valuation of Reverse Mortgage with Regular Tenure Payments,” *Insurance: Mathematics and Economics*. 51(2): 430-441.
- Li, J. S. H., M. R. Hardy & K. S. Tan
 2010 “On Pricing and Hedging the No-Negative-Equality Guarantee in Equity Release Mechanisms,” *Insurance: Mathematics and Economics*. 51(2): 430-441.
- Liaw, A. & M. Wiener
 2002 “Classification and Regression by Random Forest,” *R News*. 2(3): 18-22.
- Lipton, Z. C.
 2016 “The Mythos of Model Interpretability,” *arXiv Preprint*. arXiv:1606.03490v3.
- Rather, A. M., Agarwal, A. & Sastry, V. N.
 2015 “Recurrent Neural Network and a Hybrid Model for Prediction of Stock Returns,” *Expert Systems with Applications*. 42: 3234-3241.
- Singla, P., M. Duhan & S. Saroha
 2022 “Different Normalization Techniques as Data Preprocessing for One Step Ahead Forecasting of Solar Global Horizontal Irradiance,” in *Artificial Intelligence for Renewable Energy Systems*. 209-230. ed. A. K. Dubey, S. K. Narang, A. L. Srivastav, A. Kumar, & V. García-Díaz, Cambridge: Woodhead.

Sirmans, G. S., D. A. Macpherson & E. N. Zietz

2005 “The Composition of Hedonic Pricing Models,” *Journal of Real Estate Literature*. 13(1): 3-43.

Strobelt, H., S. Gehrmann, B. Huber, H. Pfister & A. M. Rush

2017 “LSTMVis: A Tool for Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks,” *arXiv Preprint*. arXiv:1606.07461v2.

Sutskever, I., O. Vinyals & Q. V. Le

2014 “Sequence to Sequence Learning with Neural Networks,” *arXiv Preprint*. arXiv:1409.3215v3.

Tekouabou1, S. C. K., Ş. C. Gherghina, E. D. Kameni, Y. Filali & K. I. Gartoumi

2024 “AI-Based on Machine Learning Methods for Urban Real Estate Prediction: A Systematic Survey,” *Archives of Computational Methods in Engineering*. 31: 1079-1095.

Vonlanthen, J.

2023 “Interest Rates and Real Estate Prices: A Panel Study,” *Swiss Journal of Economics and Statistics*. 159, 6.

Wandhe, A., L. Sehgal, H. Sumra, A. Choudhary & M. Dhone

2023 “Real Estate Prediction System Using ML,” *11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP)*.

Wing, C. K. & T. L. Chin

2003 “A Critical Review of Literature on the Hedonic Price Model,” *International Journal for Housing Science and Its Applications*. 27(2): 145-165.

Vinyals, O., A. Toshev, S. Bengio & D. Erhan

2015 “Show and Tell: A Neural Image Caption Generator,” *IEEE Conference on Computer Vision and Pattern Recognition*.

Yilmaz, B. & T. S. Selcuk-Kestel

2018 “A Stochastic Approach to Model Housing Markets: The US Housing Market Case,” *Numerical Algebra*. 8(4): 481-492.

Zakaria, F. & F. A. Fatine

2021 “Towards the Hedonic Modelling and Determinants of Real Estates Price in Morocco,” *Social Sciences & Humanities Open*. 4, 100176.