天然氣價格預測與市場決策支援：基於決策樹的實證分析

Natural Gas Price Forecasting and Market Decision Support: An Empirical Analysis Based on Decision Trees

陳昆皇[1*]蔣昌哲[2]林俊雄[1]陳思翰[1]

[1]明志科技大學工業工程與管理系

[2]明志科技大學工業人工智慧學士學位學程

[*]kunhuang@mail.mcut.edu.tw

摘要

天然氣為全球能源體系之關鍵能源，其價格變動直接影響能源供需結構、產業成本配置與政策規劃方向。隨著天然氣於發電、工業製程及民生部門之使用比重持續上升，價格波動已超越市場層次，成為企業風險管理、家庭支出安排與總體經濟穩定的重要變數。因此，建構具高準確性與穩定性的價格預測模型，對能源市場決策與風險控管具有實質意義。既有天然氣價格研究多以傳統時間序列與計量模型分析價格趨勢與季節循環，此類方法普遍依賴線性假設與歷史價格延伸，對氣候變動、庫存調整與地緣事件等因素所引發之結構轉折反應有限，因而限制預測效能。為克服上述不足，本研究蒐集 29 項原始市場變數，並透過資料前處理與時間特徵等多元變數，使用決策樹模型在訓練集（2020 年以前）之決定係數 R^2 達 0.9769，能精準學習歷史市場之定價規律。然而，在面臨 2020 年後因 COVID-19 疫情與俄烏戰爭等引發之市場結構性轉變，測試集之預測誤差顯著擴大（R^2 降至 0.3743），導致預測能力受限。研究結果指出，整合市場動態資訊，可有效提升模型對價格變動與季節結構之掌握能力，但面對疫情與俄烏戰爭等引發之市場結構性轉變，需要進一步強化預設模型。

關鍵詞：決策樹、能源價格預測、多變數建模、天然氣市場、決策支援系統

Abstract

Natural gas is a key energy source in the global energy system, and its price fluctuations directly influence energy supply and demand structures, industrial cost allocation, and policy planning. As the use of natural gas continues to increase in power generation, industrial processes, and residential sectors, price volatility has extended beyond the market level and has become an important factor affecting corporate risk management, household expenditure planning, and overall macroeconomic stability. Therefore, constructing forecasting models with high accuracy and stability is substantially important for decision-making and risk control in energy markets. Existing studies on natural gas prices have primarily employed traditional time-series and econometric models to analyze price trends and seasonal cycles. These approaches generally rely on linear assumptions and historical price extrapolation, resulting in limited responsiveness to structural changes caused by climate variability, inventory adjustments, and geopolitical events, thereby constraining the forecasting performance. To address these limitations, this study collected 29 original market variables and incorporated multiple factors, including data preprocessing and temporal features. Using a decision tree model, the training set (before 2020) achieved a coefficient of determination (R2) of 0.9769, demonstrating a strong capability to learn historical market pricing patterns. However, when confronted with structural market changes after 2020 caused by the COVID-19 pandemic

and the Russia–Ukraine war, the prediction error in the testing set increases substantially (R2 decreases to 0.3743), resulting in limited forecasting capability. The findings indicate that integrating market dynamic information can effectively improve the model's ability to capture price fluctuations and seasonal patterns. Nevertheless, under structural market changes induced by events such as the pandemic and the Russia–Ukraine war, further enhancement of the forecasting framework is required.

Keywords: Decision Tree; energy price forecasting; multivariate modeling; natural gas market; decision support system.

1.緒論

在淨零轉型與能源結構調整之趨勢下，天然氣逐漸成為連結傳統化石燃料與再生能源之過渡性能源。其具備良好調節能力、穩定供能特性及相對較低之碳排放，使其於電力系統與製造部門中占有關鍵地位(Ember, 2024)。隨著天然氣需求持續增加，天然氣價格波動已不僅侷限於能源市場，亦影響發電成本、產業競爭力與家庭能源支出，並可能透過成本傳遞影響整體物價與經濟穩定。因此，建立具準確性與穩健性的天然氣價格預測模型，已成為企業風險管理與政策規劃之基礎。

既有天然氣價格研究多以傳統時間序列與計量模型分析價格行為，例如 ARIMA 與 SARIMA 用以描述趨勢與季節循環(Contreras et al., 2003)。然而，此類方法主要依賴線性假設與單一價格序列，對政策調整、極端氣候(Chen et al., 2023)及地緣事件(Kilian & Zhou, 2022)所造成之結構變化反應有限，於市場波動加劇或制度轉換情境下預測能力易顯著下降。此外，多數研究未能充分納入庫存水位與總體經濟變數等因素，使模型難以呈現供需互動與價格形成過程，進而限制解釋力與實務應用價值。如何整合多元資訊並提升模型於複雜環境下之預測能力，遂成為天然氣價格預測的重要課題。

為回應上述不足，本研究導入決策樹方法，建構以供需基本面、季節影響與時間特徵之預測架構，期望提升天然氣售價預測準確度與穩定性，並將預測結果應用於採購規劃與風險評估。相較於僅依賴歷史價格之模型，所提出之多變數方法能更完整解釋價格變動來源，並為能源市場決策提供具依據性的分析工具。

2.文獻

2.1 天然氣市場與價格影響因素

為建構天然氣價格預測模型，本節從供需基本面、季節影響與時間特徵三個面向梳理價格形成機制。

2.1.1 供需基本面

與全球高度整合的原油市場不同，天然氣長期受限於管線與液化天然氣基礎設施，呈現明顯區域定價特性。其價格多由各區域市場之內部供需條件決定，缺乏如原油般全球統一之定價基準，呈現市場分割現象。然而，隨著液化天然氣儲運技術的突破，以及跨國能源貿易的蓬勃發展，各區域市場間的價格連動逐漸增強，市場整合程度提升(Shi & Variam, 2021)。此一轉變使天然氣價格同時反映區域條件與國際能源動態，增加價格形成供給和需求機制的複雜度。

在供給面，技術進步與基礎設施能力對價格具有長期影響。頁岩氣開採技術的突破提升北美產量，改變全球供需格局並影響價格水準(Zhang et al., 2020；Kilian & Zhou, 2022)。此外，管線容量與儲氣設施效率決定供應彈性；一旦出現地緣衝突或運輸中斷，供應受限將迅速放大價格波動。

在需求面，天然氣消費與經濟活動、發電需求及氣候條件密切相關。隨著工業與電力需求成長，天然氣在能源組合中的角色持續擴大(Ember, 2024)。替代能源價格，特別是原油價格，也會透

過替代效果影響天然氣需求。研究指出，原油與天然氣價格存在長期關聯與動態互動(Villar et al., 2006)，顯示兩者在能源市場中具有顯著連動性。

除基本面因素外，政策制度亦為重要外生變數。碳定價與排放交易制度改變燃料成本結構，進而影響天然氣需求。例如碳價上升會提高燃煤成本，使天然氣發電比重增加(Villar & Joutz, 2006)。再生能源政策亦可能改變中長期需求趨勢。

綜合上述因素，天然氣價格同時受到供需條件、技術變化、替代能源與政策環境影響，呈現多因素交互作用特性，難以完整反映此一複雜機制。因此，本研究將儲存量、消耗量與相關供需指標納入模型，以強化價格預測能力。

2.1.2 季節影響

在各類影響因素中，季節性需求為天然氣價格最具規律性的來源之一。由於天然氣廣泛用於供暖與發電，其需求對氣溫變化高度敏感，尤其在溫帶與寒帶地區呈現明顯年度循環。冬季低溫推升供暖需求並壓縮庫存，價格隨之上升；春秋需求減弱，價格通常回落(Mu et al., 2007)。此種週期性構成天然氣價格的重要結構特徵。

若預測模型未能辨識此類週期模式，將產生系統性偏差並降低準確度(Abdollahi, 2020)。近年氣候變遷與極端天氣事件增加，使傳統季節規律出現擾動。寒潮與熱浪可能同時影響需求與供應，進一步放大價格波動。研究指出，極端氣候會提高能源市場波動程度與風險(Chen et al., 2023)。

為量化氣候影響，文獻普遍使用供暖度日(Heating Degree Days, HDD)與冷卻度日(Cooling Degree Days, CDD)作為溫度指標，以衡量能源消耗強度。過去研究顯示，納入氣象變數可提升價格變動的解釋能力與預測表現(Chen et al., 2023)。近期研究亦利用季節資料捕捉極端氣溫對價格的即時影響。

因此，氣候與季節因素為天然氣價格短期變動的重要來源。本研究據此將相關溫度指標納入模型，以提升對季節循環與極端氣候衝擊的辨識能力。

2.1.3 時間特徵

除供需基本面與季節因素外，時間本身亦蘊含重要資訊，對天然氣價格形成與預測具有關鍵意義。能源價格通常呈現趨勢性與循環性等時間結構特徵，使其變動不僅受當期基本面影響，亦與過去價格行為密切相關(Hamilton, 2020; Box et al., 2015)。因此，適當建構時間特徵，有助於模型掌握價格動態調整過程與市場慣性。

價格波動往往呈現異質變異現象，即高波動期間集中出現，低波動期間亦相對平穩。此種波動叢聚特性已廣泛存在於能源與金融市場。若模型未納入反映波動程度的時間特徵，將難以捕捉風險變化與不確定性擴散過程。因此，除價格水準外，亦可透過移動標準差、歷史波動率或區間震盪幅度等指標衡量時間風險特徵。

長期趨勢亦為重要時間結構來源。隨著能源轉型、技術進步與政策調整，天然氣在能源組合中的角色持續演變，可能形成結構性上升或下降趨勢。若未考量時間趨勢變數，模型可能誤將結構變化解釋為短期波動，降低預測穩定性。因此，在特徵設計上，可透過滾動平均或累積變動率等方式，刻畫長期趨勢效果(Weron et al., 2014)。

2.2 預測模型

隨著資料處理能力與分析技術的提升，能源價格預測方法已由傳統統計模型逐步轉向廣泛的機器學習領域。相較早期著重時間序列內在結構的計量模型，近年研究更強調整合多元變數、處理非線性關係，以及提升預測能力，以因應能源市場波動加劇與制度變化頻繁的特性。本節先回顧傳統時間序列模型的理論基礎與限制，再說明決策樹方法在能源價格預測中的優勢。

2.2.1 傳統統計模型

早期天然氣價格預測多採用自迴歸整合移動平均模型及其季節性延伸模型(Contreras et al., 2003)。此類方法透過自相關結構與季節循環，能有效描述價格序列的趨勢與週期變化，且模型形式明確，參數具經濟解釋性(Assimakopoulos et al., 2000)。

然而，這類模型多建立在線性關係與價格平穩假設之上，且主要依賴歷史價格資訊進行推估。當市場受到政策調整、供需失衡或極端氣候等外生因素影響時，價格可能出現結構性變化，導致模型參數失效與預測誤差擴大。此外，傳統時間序列方法難以同時整合變數間的交互作用，使其在複雜市場環境下的因子解釋力與模型預測能力逐漸受限。因此，難以滿足當前能源價格預測對準確度與穩定性的要求。

2.2.2 決策樹模型

為克服上述限制，決策樹方法逐漸被應用於能源價格預測。與傳統計量模型相比，決策樹不需事先設定特定函數形式，能透過資料自動學習變數間的複雜關係(Villar & Joutz, 2006)。此特性使其更適合處理非平穩、波動性高且受多重因素影響的能源市場資料。

透過特徵工程，研究者可將供需基本面、季節影響與時間特徵等資訊轉換為多維輸入特徵，使價格預測問題轉化為監督式學習任務(Lago et al., 2018)。此一作法有助於提升模型預測表現。

在能源價格預測研究中，機器學習模型已被廣泛應用。相關研究指出，機器學習方法透過非線性函數近似能力，可有效處理多變數與高度波動之市場資料。其中，集成式方法藉由多模型組合降低偏誤與變異數，在非線性環境下通常表現較為穩定。以隨機森林為例，其透過多棵決策樹進行重抽樣與隨機特徵選擇，能降低過度擬合風險並捕捉變數間交互關係(Breiman, 2001)。此外，該方法所提供之特徵重要性衡量，亦有助於辨識影響能源價格之關鍵因子，使模型兼具預測能力。

基於上述考量，本研究方法採用單一決策樹模型（Decision Tree Model）。有別於傳統線性迴歸需依賴嚴格的線性與常態分配假設，決策樹能靈活捕捉各解釋變數對天然氣價格之非線性影響，它可將複雜的定價機制轉化為直觀的樹狀分支結構（例如：當庫存低於特定水準，且適逢冬季高用量時的價格反應路徑），並透過特徵重要性分數，客觀量化各變數對價格變動的貢獻度。配合適當的參數設定的同時，適度控管過度擬合所造成之樣本外不穩定問題。因此，本研究以決策樹作為主要預測與分析工具，以提升研究結論之實務證明。

2.3 居民天然氣價格作為預測對象

天然氣價格可依使用部門區分為居民與工業等類型，不同部門受合約結構、需求彈性與政策規範之影響，其價格形成機制與波動特性有所差異。本研究選擇居民天然氣價格作為預測對象，主要基於資料穩定性與研究代表性兩項考量。居民天然氣價格資料，具有較高的完整性與連續性，適合進行長期時間分析。其波動幅度較為平穩，較能反映整體供需變動趨勢。本研究以居民天然氣價格為核心變數，作為預測分析之主要對象。

2.4 小結

綜合前述文獻可知，天然氣價格的形成並非單一因素所決定，而是供需基本面、季節規律與時間動態共同作用的結果。供需變數反映市場結構與庫存調節能力，是價格變動的核心基礎；季節與氣候因素則構成穩定且可重複的年度循環，對短期波動具有明顯影響；而時間特徵則揭示價格的延續性、波動聚集與長期趨勢，使價格變動呈現動態演進過程。三者彼此交織，使天然氣市場具有高度複雜性與不確定性。在預測方法上，傳統單變量時間序列模型（如 ARIMA 類模型）雖能描述價格的自我延續特性與短期波動結構，但其主要依賴歷史價格資訊，對於多重外生變數之整合能力有限，特別是在能源市場受到供需基本面、國際原油價格、氣候條件與總體經濟變數共同影響的情境下，單一時間序列架構往往難以完整反映價格形成機制。

因此，本研究將天然氣價格視為多重經濟與市場因素交互作用下之結果，採用單一決策樹模型建構預測架構。透過將供需基本面變數(X1~X27)、季節性虛擬變數(X28~X29)與時間趨勢(X30~X43)項納入同一模型框架中，亦能透過「特徵重要性（Feature Importance）」指標，精準量化不同因素在價格變動中之相對貢獻度。此種建模方式能將複雜的市場定價機制具象化為清晰的決策路徑，同時亦有助於政策與市場決策應用。

3.研究方法

本研究以美國能源資訊署公布之月度能源統計資料為主要資料來源。該資料涵蓋天然氣生產、進出口、庫存、消費與價格等供需資訊，具長期一致性與完整性，廣泛應用於能源經濟研究。採用官方資料可確保變數定義明確，並提升研究結果之可驗證性與實務參考價值。

本研究以居民天然氣售價為預測目標變數，整合供需基本面、季節影響與時間特徵，共建構 43 變數。供需基本面內容包括天然氣地下儲存量與天然氣總消耗量，季節影響內容包括 HDD 與 CDD，而時間特徵則是所有變數資料做處理，上述變數可反映市場供需狀態與需求強度，其定義與資料說明列於表 1。

表 1. 變數

X1	天然氣總開採量	X2	天然氣井開採量	X3	油井天然氣開採量
X4	頁岩氣開採量	X5	煤層氣井開採量	X6	天然氣銷售生產量
X7	天然氣液體產出量	X8	乾天然氣產量	X9	天然氣進口量
X10	天然氣管線進口量	X11	液化天然氣進口量	X12	天然氣出口量
X13	天然氣管線出口量	X14	液化天然氣出口量	X15 (供需基本面)	天然氣地下儲存總量
X16	地下儲氣庫基礎氣量	X17	地下儲氣庫可運用氣量	X18	注入地下儲氣庫的天然氣量
X19	地下儲氣庫提取量	X20	地下儲氣庫淨提取量	X21 (供需基本面)	天然氣總消耗量
X22	天然氣對消費者供應量	X23	住宅天然氣消耗量	X24	商業消費者天然氣供應量
X25	工業天然氣消耗量	X26	車用天然氣燃料消耗量	X27	電力部門天然氣供應量
X28(季節影響)	HDD(供暖度)	X29(季節影響)	CDD(冷卻度)	X30	time_trend
X31	month_2	X32	month_3	X33	month 4
X34	month_5	X35	month_6	X36	month 7
X37	month_8	X38	month_9	X39	month 10
X40	month_11	X41	month_12	X42	落後 1 期價格 (price_lag_1)
X43	落後 12 期價格 (price_lag_12)	Y	居民天然氣價格		

3.1 資料前處理

本研究以美國能源資訊署公布之能源統計資料為主要來源，研究期間自 2008 年 01 月起至 2022 年 12 月止。原始資料涵蓋天然氣生產、進出口、庫存、消費與價格等供需資訊。為確保多源資料頻率之一致性，本研究於資料前處理階段採用月度重採樣的方式進行彙總，處理後之總樣本數為 180 筆。

在模型訓練與驗證架構上，因天然氣價格等時間序列資料具有高度的時間相依性，為避免未來資訊外洩並符合實務預測情境，本研究嚴格遵循時間順序進行樣本切分，而非採用隨機抽樣。整體樣本以 2020 年 1 月 1 日為基準點切分為訓練集與測試集，切分比例精準落在 80:20（訓練集 144 筆，測試集 36 筆）。2020 年 1 月之前之歷史資料作為訓練集，用以學習價格特徵與擬合模型參數，2020 年 1 月起之資料則作為樣本外測試集，以客觀評估模型之預測效能與泛化能力。

3.2 時間特徵與模型輸入變數

為使模型能同時掌握市場供需狀態、季節規律與價格動態，本研究依經濟意涵與資訊來源，將輸入變數分為三類：供需基本面、季節影響與時間特徵。此種分層設計可同時整合外部供需資訊與內部時間動態，使模型在預測時具備較完整的市場背景。

第一供需基本面變數用以反映天然氣市場的供需平衡與庫存調節能力。價格在本質上源於供需互動，因此納入相關指標有助於模型直接掌握市場壓力與供應彈性。本研究選取地下儲存量與總消耗量等原始市場變數。儲存量可衡量供應緩衝能力，當庫存偏低時，市場供應緊張，價格上行風險增加；消耗量則反映需求強度，其增加通常

伴隨價格上升。上述資料皆以月度建構，並進行季節調整，以降低固定季節效應對估計結果的干擾，使模型更能辨識實質供需變化。

第二季節影響用以捕捉天然氣需求隨氣候變動所產生的規律模式。天然氣廣泛應用於供暖與發電，因此需求對氣溫變化相當敏感，特別是在冬季與夏季用電高峰期間，價格常出現明顯波動。為反映此類週期性特徵，本研究納入月份變數，以區分不同月份的需求結構差異，使模型能辨識年度循環所帶來的價格變化。透過季節性指標的設定，可避免模型將固定的季節規律誤判為隨機波動，進而提升預測穩定性。

第三時間特徵用以描述長期趨勢與年度循環。天然氣需求受氣候與使用習慣影響，呈現穩定的季節模式。若未將時間資訊納入模型，容易忽略此類規律而產生預測偏差。因此在進行預測時，必須考量歷史同期資料的參考價值。例如在預測某年度 1 月價格時，模型會同時納入前一期價格以及一年前 1 月的相關資訊，使其能掌握年度重複模式與價格延續效果。透過此設計有助於提升模型對週期變動的辨識能力，並降低季節錯置所造成的誤差。

綜合而言，供需基本面、季節與時間三類特徵分別對應供需結構、季節規律與歷史依存關係。此一整合式設計可同時反映長期趨勢與短期變動，為後續決策樹模型訓練與預測分析提供完整的資訊基礎。

3.3 決策樹模型

為評估因子對預測效能的影響，本研究建立決策樹模型作為方法。透過對特徵空間進行分割，將複雜的市場變數與價格關係轉化為具備邏輯層次的樹狀決策路徑。相較於傳統線性模型，決策樹不需預設目標變數與解釋變數間之函數形式，能更靈活地捕捉天然氣價格。本研究以美國居民天然氣售價為應變數，其目標在於找尋最佳的特徵與切分點，使得分割後之子節點內的價格預測準確。

決策樹模型能提供定價機制的初步分析基礎。此種建模亦能透過特徵重要性分析（Feature Importance），協助辨識驅動天然氣價格之核心關鍵因子。

3.4 模型評估指標

為全面比較各模型之樣本外預測能力，本研究採用多指標評估架構，從絕對誤差、比例誤差與尺度標準化誤差三個面向衡量模型表現。單一指標往往偏重特定誤差特性，可能產生評估偏誤；因此透過多元指標並行，可同時檢視預測準確度、穩定性與相對優勢，提供更完整的效能判斷基礎。

3.4.1 模型評估指標：決定係數

決定係數用以衡量模型對價格變異的解釋程度，反映模型相較於以樣本平均值預測所能降低的變異比例。數值越接近 1，表示模型越能掌握價格波動結構。與誤差指標不同，R^2 著重整體擬合能力，可用以判斷模型是否有效捕捉市場動態。R^2的計算公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (4)$$

3.4.2 模型評估指標：平均絕對比例誤差

為評估模型是否真正優於簡易基準預測(naïve forecast)，本研究採用平均絕對比例誤差（Mean Absolute Scaled Error, MASE）。相較前述誤差度量僅提供絕對或比例誤差大小，MASE 透過與季節性樸素模型進行相對比較，可量化模型之「預測技能(forecast skill)」。

當 $MASE < 1$ 時，代表模型表現優於基準方法，具有實質預測價值；反之，若 $MASE \geq 1$，則表示模型未能超越簡單歷史延續策略。此種相對評估機制在預測競賽與時間序列文獻中被廣泛視為公平且具可比性的衡量標準。然而，若預測區間發生結構性轉折或遭遇極端外部衝擊（如俄烏戰爭），因測試集之整體波動幅度與不確定性遠大於訓練集，可能導致所有模型之 MASE 值皆大於 1。在此種極端非平穩情境下，MASE 的絕對基準意義將減弱，應以不同預測模型間 MASE 數值的相對大小，作為評估模型優劣之主要依據。其中分

母項代表「季節性樸素模型在訓練集上之誤差。T 為訓練集之樣本總數，m 為資料之季節週期長度（本研究採用月度資料，故設定 m＝12），$\mathbf{Y_t}$ 與 $\mathbf{Y_{t-m}}$ 分別為歷史訓練集中第 t 期與落後 m 期之實際價格。

$$\mathrm{MASE} = \frac{\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i|}{\frac{1}{T-m}\sum_{t=m+1}^{T}|Y_t-Y_{t-m}|} \qquad (7)$$

3.4.3 模型評估指標：均方根比例誤差

為同時考量相對誤差與極端偏差影響，本研究進一步採用均方根比例誤差（Root Mean Squared Scaled Error, RMSSE）。該指標保留平方懲罰特性，並以季節性基準模型為比較基準，適合評估模型在高波動情境下的穩健程度。RMSSE 的計算公式如下：

$$\mathrm{RMSSE} = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\frac{1}{T-m}\sum_{t=m+1}^{T}(Y_t-Y_{t-m})^2}} \qquad (8)$$

綜合而言， R^2 評估整體解釋力，而 MASE 與 RMSSE 則判斷模型是否優於簡易預測方法。透過此多面向評估架構，可避免單一指標造成判斷偏誤，並為後續模型比較提供一致且客觀的依據。

4.研究結果

本研究將資料分成訓練集和測試集，表 2 呈現了決策樹模型之實證結果。實證數據顯示，決策樹在訓練集（2020 年以前）展現出極為優異的配適能力，其決定係數（R^2）高達 0.9769，且反映預測技能之 MASE（0.2453）與 RMSSE（0.2760）皆遠小於 1。此結果證明模型已成功學習天然氣市場供需基本面與價格間之定價規律。

然而，在進入 2020 年 1 月起之樣本外測試集後，模型績效出現顯著衰退。測試集之 R^2 降至 0.3743。更值得注意的是，測試集之 MASE（2.0328）與 RMSSE（2.2242）皆顯著大於 1。

針對測試集績效落後於訓練集之現象，本研究認為主要源於測試集區間所發生之極端外生衝擊(COVID-19 (Nyga-Łukaszewska and Aruga, 2020)與俄烏戰爭 (Tsuji, 2024)等)導致市場產生結構性轉變。此種非規律性波動完全脫離了模型在訓練集所學得之常態定價模式，導致天然氣價格脫離區域供需基本面而發生非理性飆漲。

表 2. 決策樹模型

訓練/測試	MASE	RMSSE	R^2
Train	0.2453	0.2760	0.9769
Test	2.0328	2.2242	0.3743

為進一步釐清測試誤差之來源，本研究採用擴展視窗（expanding-window）方式進行時間切點之敏感度分析，於原有 2007–2019 年訓練集之外，將訓練期間依序延伸至 2020 年與 2021 年，並以其後之期間作為測試集，結果彙整如表 3。就訓練集而言，三種切分方式之績效高度一致，R^2 介於 0.9750 至 0.9782，MASE 介於 0.2453 至 0.2623，顯示模型之樣本內配適能力不因訓練期間延伸而改變。然而，樣本外預測能力並未隨近期資料之納入而改善：當訓練集延伸至 2020 年時，測試集（2021–2022 年）之 R^2 由 0.3743 降至 0.0863，MASE 與 RMSSE 分別升至 2.8254 與 2.7005；即使再將 2021 年納入訓練，2022 年測試集之 R^2 僅回升至 0.3194，MASE（2.4937）與 RMSSE（2.3658）仍遠大於 1。

表 3. 擴展視窗

切分	訓練/測試	MASE	RMSSE	R^2
1	Train (2007~2019)	0.2453	0.2760	0.9769
	Test (2020~2022)	2.0328	2.2242	0.3743
2	Train (2007~2020)	0.246	0.2659	0.9782
	Test (2021~2022)	2.8254	2.7005	0.0863
3	Train (2007~2021)	0.2623	0.2899	0.9750
	Test (2022~2022)	2.4937	2.3658	0.3194

上述結果具有兩層意涵。其一，將 2020 年自測試集移除後，各項誤差指標不降反升，顯示預測誤差主要集中於 2021–2022 年，與前述極端外生衝擊導致市場結構性轉變之推論相符。其二，即使訓練集已涵蓋 2021 年之價格上漲階段，模型對 2022 年之預測能力仍未明顯恢復，表示此一誤差無法僅藉由增加近期訓練樣本加以消除。本研究認

為，此現象與決策樹模型之預測值受限於訓練集目標變數範圍、無法外推之特性有關：當測試期間價格水準超出歷史區間時，模型即產生系統性低估。因此，目前較大之測試誤差應為市場結構性轉變與決策樹模型外推能力受限兩者共同作用之結果，而非單純源於訓練資料不足。惟三組測試集皆涵蓋 2021–2022 年之極端期間，模型於常態市場下之樣本外泛化能力仍有待後續研究進一步驗證。

由於決策樹模型係基於歷史數據進行特徵切分，當遭遇從未出現過之極端價格極值時，模型容易過度記憶歷史雜訊，進而導致樣本外泛化能力受限。此一實證發現具體揭示了單一機器學習演算法在應對地緣政治與全球性公衛危機時之限制，亦說明在波動性極高之能源市場中，導入動態監控與模型重新訓練機制之必要性。雖然決策樹在測試集預測上受外生衝擊干擾，但其仍透過基於決策樹中之平均不純度減少量（Mean Decrease in Impurity, MDI）演算法計算而得。由排行結果可知（如表 4 所示），除了落後價格資訊外，天然氣總消耗量與輸送量等基本面變數仍佔據重要權重。此外，時間趨勢項與特定月份變數亦提供了顯著的資訊增益，證實本研究所建構之「資料前處理與動態變數」確實能協助模型掌握價格之長、短期變動特徵。

表 4. 前 5 名重要性

排名	變數名稱	**重要性**
1	美國商業用戶天然氣輸送量 (Natural Gas Deliveries to Commercial Consumers)	0.732274
2	落後 12 期價格 (price_lag_12)	0.155808
3	落後 1 期價格 (price_lag_1)	0.047849
4	天然氣總消耗量 (U.S. Natural Gas Total Consumption (MMcf))	0.016916
5	天然氣輸送量 (Natural Gas Delivered to Consumers in the U.S. (MMcf))	0.009115

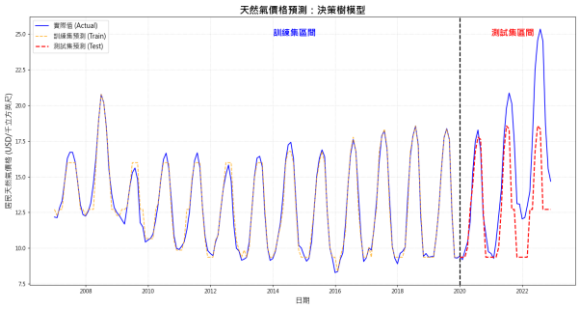


圖 1. 決策樹模型

5.結論

本研究建立了以供需基本面、季節影響與時間特徵多元變數為基礎之天然氣價格預測模型，並導入具備非線性捕捉能力之決策樹演算法進行實證分析。研究結果顯示，決策樹模型在訓練集展現出卓越之配適度（R^2 達 0.9769），能精準學習歷史市場規律；但在面臨 2020 年後之樣本外測試集時，預測效能出現顯著衰退（R^2 降至 0.3743），且 MASE 與 RMSSE 指標皆大於 1。

此實證結果具體指出了天然氣是一個極易受外生衝擊干擾的系統。模型在測試集之失準，客觀反映出疫情與俄烏戰爭等事件引發了市場之結構性轉變。在方法上，本研究透過特徵工程，使模型能同時辨識供需基本面、季節影響與時間特徵等結構訊號。相較傳統統計模型主要依賴自相關結構，本研究之建模方式能直接納入市場狀態資訊，因而在趨勢轉折與市場波動時仍能維持穩定預測能力，顯示資料整合對能源價格建模具有關鍵作用。

在應用面，所建構之模型可作為能源決策分析的輔助工具。較準確的價格預測有助於政府進行能源供應規劃與庫存管理，也可協助企業判斷採購時機與控制價格風險，降低能源成本波動對營運的影響。

本研究雖透過決策樹模型釐清了價格驅動因子。未來研究可進一步嘗試其他模型等資料，以期在極端市場波動下提升模型預測能力。此外，如何將地緣政治風險與全球供應鏈指標量化並納入預測特徵，亦為能源資料分析值得深入探討之重要方向。

6.參考文獻

1. Abdollahi, H., & Ebrahimi, S. B. (2020). A new hybrid model for forecasting Brent crude oil price. *Energy*, *200*, 117520.
2. Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting. *International Journal of Forecasting, 16*(4), 521–530.
3. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.
4. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
5. Chen, Y., Hartley, P. R., & Lan, Y. (2023). Temperature, storage, and natural gas futures prices. Journal of Futures Markets, 43, 549–575.
6. Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems, 18*(3), 1014–1020.
7. Ember, E. I. (2024). *Statistical Review of World Energy (2024), Energy Institute*.
8. Hamilton, J. D. (2020). *Time series analysis*. Princeton University Press.
9. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.
10. Kilian, L., & Zhou, X. (2022). Oil prices, exchange rates and interest rates. *Macroeconomic Dynamics, 22*(8), 1929–1972.
11. Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research, 259*(2), 689–702.
12. Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison. *Applied Energy, 221*, 386–405.
13. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE, 13*(3), e0194889.
14. Mu, X. (2007). Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics*, 29(1), 46-63.
15. Nyga-Łukaszewska, H., & Aruga, K. (2020). Energy prices and COVID-immunity: The case of crude oil and natural gas prices in the US and Japan. *Energies*, *13*(23), 6300.
16. Serletis, A., & Herbert, J. (1999). The message in North American energy prices. *Energy Economics, 21*(5), 471–483.
17. Shumway, R. H., & Stoffer, D. S. (2006). *Time series analysis and its applications: With R examples* (2nd ed.). Springer.
18. Tsuji, C. (2024). Simmering tensions on the Russia–Ukraine border and natural gas futures prices: identifying the impact using new hybrid GARCH. *Energy Informatics*, *7*(1), 37.
19. Villar, J. A., & Joutz, F. L. (2006). The relationship between crude oil and natural gas prices (EIA Working Paper). *Energy Information Administration*.
20. Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), 1030-1081.
21. Zhang, W., & Hamori, S. (2020). Do machine learning techniques and dynamic methods help forecast US natural gas crises?. *Energies*, *13*(9), 2371.