

財務工程專題期末報告

結合大離差理論與機器學習之預測風險指標建構

110071025 計財大四 張智傑

一、研究摘要

本研究聚焦於金融市場中極端下跌事件的預測與風險判別機制建構。鑑於傳統避險策略常因缺乏有效的啟動時機判斷而導致報酬侵蝕或避險失靈，本文嘗試導入大離差理論與市場統計特徵，建構具備前瞻性的風險預警指標，協助投資者在市場即將出現系統性風險時進行策略調整。

本研究首先提出一組以報酬歷史行為為基礎的風險判別指標組合，包含：大離差理論所推導之 Rate Function、Z-score、Skewness、Kurtosis、絕對動能與歷史 VaR 等共五項統計指標。接著使用 Logistic Regression 模型，以 2015 年至 2019 年資料為訓練集，2020 年至 2025 年資料為測試集，進行極端下跌（定義為單日跌幅超過 -5%）事件的預測。

實證結果顯示，模型具備穩定的預測能力，並能有效識別過半數以上的極端下跌事件，AUC 與混淆矩陣指標亦顯示其在分類上具備一定準確度，為後續動態資產配置與風險控制提供量化依據。

二、研究動機與目的

在金融市場中，極端事件雖然發生機率偏低，卻常造成資產價格劇烈波動，對投資組合帶來非對稱且重大的負面衝擊。這類事件多屬於系統性風險，難以透過傳統的資產配置與分散策略有效迴避，亦使得風險控管成為資產管理中的核心課題。

為了降低此類風險對投資組合的衝擊，市場參與者常透過衍生性金融商品（如期貨與選擇權）進行避險操作。然而，若無法正確掌握極端事件的發生時機，過度提前或長期持有避險部位反而可能侵蝕整體投資報酬，使避險行為本身成為績效拖累的來源。

因此，若能建立一套具備前瞻性、可預警極端市場下跌風險的預測機制，協助投資者在適當時機啟動風險控管行動，將有助於提升避險效率，同時保留市場大多數時間的報酬潛力。

本研究即致力於此一目標，嘗試結合大離差理論（Large Deviation Principle, LDP）與市場統計特徵，建構反映極端風險機率的預測指標，並透過機器學習模型進行驗證與應用。此機制未來亦可延伸為選擇權或其他策略調整的依據，進一步落實風險導向型資產配置的核心精神。

三、文獻回顧與市場分析

3.1 文獻回顧

過去對於極端市場事件的預測，多聚焦於風險量化方法的建構與改良，例如歷史模擬法（Historical Simulation）、蒙地卡羅模擬（Monte Carlo Simulation）、GARCH 類波動模型，以及 Value at Risk（VaR）等尾部風險衡量工具。然而，這些傳統方法大多依賴歷史報酬分布的穩定性與假設條件，對於突發性、非對稱性事件的反應仍具侷限。

為突破上述限制，部分研究導入極端理論（Extreme Value Theory）與大離差理論（Large Deviation Theory, LDT）來捕捉稀有但破壞力強的事件機率。其中，大離差理論提供了計算極小機率事件所需的指數收斂率（rate function）架構，已廣泛應用於統計物理、排隊理論、風險管理等領域。近期有學者（如 Glasserman, 2004；Dembo & Zeitouni, 1998）亦嘗試將 LDT 應用於金融市場，作為市場異常偏離的風險預警工具。

此外，隨著機器學習技術在金融市場中的快速發展，多篇文獻指出：以統計特徵為輸入的分類模型（如 Logistic Regression、Random Forests、XGBoost）能有效提升極端事件辨識的精準度，尤其在非線性與高維風險特徵共存的情境下更具優勢。

3.2 市場背景與問題意識

以 2020 年以來的市場為例，無論是 COVID-19 疫情爆發、美國升息循環，或是地緣政治衝突（如俄烏戰爭、台海緊張）、乃至於川普的對等關稅政策等，皆造成台灣股市與全球市場出現多次瞬間性下跌與跳空缺口。這些事件雖在發生前鮮有明確警訊，但往往造成投資人巨大資本損失。

尤其在被動型投資盛行之下，傳統 Buy and Hold 策略在極端風險面前顯得

毫無防禦能力。而現行多數風險管理機制仍偏重事後回溯，難以滿足實務中「事前判斷」的避險需求。因此，如何事前偵測潛藏風險、於市場即將下跌時提供預警，成為投資策略升級的核心命題。

四、風險指標建構

為有效預測金融市場中的極端下跌事件，本研究設計一組具備理論依據與實務解釋力的風險指標，並搭配機器學習模型進行分類預測。以下分別說明本研究在風險衡量、統計假設之方法。

4.1 風險衡量假設與理論基礎

本研究假設資產的報酬分布可近似為獨立同分布（i.i.d.）的常態分布。此假設雖不完全符合真實市場特性，但有助於建立穩定的指標計算機制，並使部分統計量（如 Z-score、Skewness、Kurtosis）具備明確解釋與數理性質。此外，透過滾動視窗計算方式，可部分緩解資料非平穩性對預測結果的影響。

4.2 指標設計與解釋

本研究共設計六項風險指標，分別從統計偏離、尾部風險與市場趨勢三大面向進行衡量，具體如下：

(1) Rate Function（LDP-based）

根據大離差理論，當隨機變數序列滿足 i.i.d. 條件時，其大幅偏離期望值的機率將以指數速度收斂。本研究以報酬的平方偏離加總對數作為近似的 rate function 形式，用以衡量當前市場狀態相對均值的異常程度。Rate Function 越大，代表發生異常報酬的「成本」越低，潛在極端事件發生機率越高。

前提假設

- 假設每日報酬 r_t 近似服從常態分布：

$$r_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$$

- 參考大型偏差理論 (Large Deviation Principle, LDP)，
我們關心隨機變數偏離其期望時，
機率衰減速度所對應的 **rate function** $I(x)$ 。

一般形式——Cramer Rate Function

$$I(x) = \sup_{\theta \in \mathbb{R}} \left\{ \theta x - \log \mathbb{E}[e^{\theta X}] \right\},$$

其中

$$\mathbb{E}[e^{\theta X}]$$

為隨機變數 X 的 mgf (moment generating function)。

套用至常態分布

對 $X \sim \mathcal{N}(\mu, \sigma^2)$ 而言，mgf 為

$$\mathbb{E}[e^{\theta X}] = \exp\left(\theta\mu + \frac{1}{2}\theta^2\sigma^2\right).$$

將之帶入Cramer Rate Function，取 \sup_{θ} 後可得

$$I(x) = \frac{(x - \mu)^2}{2\sigma^2}.$$

本研究之 Rate Function 指標

因此，對每日報酬 r_t ，令

$$\mu_t = \text{rolling_mean}_{60}(r_{t-1}, \dots, r_{t-60}), \quad \sigma_t = \text{rolling_std}_{60}(r_{t-1}, \dots, r_{t-60}),$$

則 大離差風險指標 定義為

$$\text{rate_func}_t = \frac{(r_t - \mu_t)^2}{2\sigma_t^2}.$$

rate function數值越大，代表當日報酬對歷史分布而言越罕見，
亦即潛在極端事件風險越高。

(2) Z-score

為報酬相對於其過去 60 日平均值的標準化距離，能夠保留方向資訊，捕捉報酬快速向負方向偏離的情況，常作為市場反轉或崩跌前兆的指標。

(3) Skewness (偏態)

評估報酬分布的非對稱性。若偏態為負，代表左尾肥厚，表示市場更可能出現大幅下跌。此特徵在金融市場風險管理中經常被視為「尾部風險」的訊號來源。

(4) Kurtosis (峰度)

衡量報酬分布的尖銳程度與尾端厚度。峰度越高表示報酬分布中包含更多極端值，代表市場波動集中，存在跳點風險。

(5) Absolute Momentum (絕對動能)

衡量報酬過去幾日的波動強度。當動能強勁時，市場雖可能持續單邊上漲或下跌，但也意味著潛在的回調壓力正在累積，對預警下跌具有前瞻性。

(6) VaR_5 (歷史 VaR)

根據過去 60 日報酬序列計算之 5% 分位數，用以估計未來最差情境下的一日損失水準。此為量化尾部風險的經典指標，可視為最低風險容忍邊界。

五、資料設計與模型訓練流程

本研究以台灣股票市場進行風險預測模型的訓練與測試。為模擬實際投資組合可能面臨的市場情境，本章節將說明資料來源、投資組合設計、事件標籤設定以及模型的整體訓練流程。

5.1 資料來源與報酬序列建構

本研究採用期間為 2015 年 1 月 1 日至 2025 年 5 月 31 日的歷史市場資料，資料頻率為每日 (daily)。為模擬一般投資人實務中較常見的大盤型投資配置，並同時反映多空部位的表現，設計如下模擬投資組合：

- 元大台灣 50 ETF (0050)：配置權重為 70%
- 元大台灣 50 反 1 ETF (0051)：配置權重為 30%

每日投資組合報酬率 (portfolio return) 依據上述兩檔 ETF 的加權報酬加總計算。該組合可近似台灣大盤波動趨勢，同時反映投資人可能的空頭避險部位，作為風險判斷基礎。

5.2 標籤設定：極端事件定義

為建立監督式學習模型，本研究需將每一日是否屬於「風險事件」進行標籤化 (Labeling)。具體定義如下：

- 極端下跌事件定義：當「隔日投資組合報酬率」跌幅超過 -5% 時，即視為一次極端下跌事件，標記為 1 (風險事件)；否則標記為 0 (非事件)。

- 設計理由：以 -5% 作為閾值，能有效排除日常小幅震盪，聚焦於具體的系統性風險發生情形，該門檻亦兼顧樣本數量與極端性，提升模型在少數類別識別上的穩定性。

5.3 模型方法與訓練方式

本研究選用邏輯回歸（Logistic Regression）作為基礎預測模型，屬於二元分類演算法，能根據風險指標輸出每一日發生極端事件的機率分數（probability score）。

- 輸入變數：本研究第三章所構建之六項風險指標
- 輸出變數：預測是否為「潛在風險事件日」（即隔日報酬 $\leq -5\%$ ）

模型特性與優點如下：

1. 屬於具可解釋性的線性模型，可觀察各指標對事件預測的邏輯權重方向
2. 計算效率高，適用於滾動更新與即時監控機制
3. 可作為日後擴展至非線性模型（如 XGBoost、Random Forest）的基礎比較版本

5.4 訓練與測試集劃分

為避免資料洩漏與過擬合，本研究依時間順序將資料劃分為訓練與測試兩組：

- 訓練資料集：2015/01/01 ~ 2019/12/31，用於模型參數估計與擬合
- 測試資料集：2020/01/01 ~ 2025/05/31，作為擬真投資場景，驗證模型預測效果

5.5 滾動預測流程

模型實施採每日更新預測機制（rolling forecast），即：

1. 每日根據前 60 日資料計算風險指標
2. 將該日的指標值輸入訓練完成的邏輯回歸模型
3. 預測該日是否為潛在極端事件日。

此一滾動式運作機制，能模擬實務中連續觀察與即時反應的風險管理需求。

六、效能驗證與預測結果

本章將依序分析模型之解釋力、特徵之共線性檢定與分類預測表現，評估本研究風險預測機制在實際運作情境下的可行性與穩定性。

6.1 模型估計結果與邏輯解釋力

透過邏輯回歸模型進行極端事件預測後，從圖一可觀察到整個回歸具有顯著性，然後所有風險指標皆具統計顯著性，且其估計方向符合經濟與市場邏輯預期。Rate Function、VaR 等反映「潛在尾部損失」或「異常偏離程度」的指標，對預測極端事件具有正向貢獻；而如 Z-score、偏態與峰度則揭示報酬分布形狀的異常變化，進一步輔助預測市場可能進入不穩定區間。

值得一提的是，Absolute Momentum 亦呈現顯著性，顯示短期內波動擴大的狀況，經常伴隨反轉與風險事件發生。整體而言，模型能有效整合六項風險因子，並具備邏輯一致與實務解釋力。

圖一：整體回歸分析結果

Optimization terminated successfully.
Current function value: 0.162947
Iterations 10

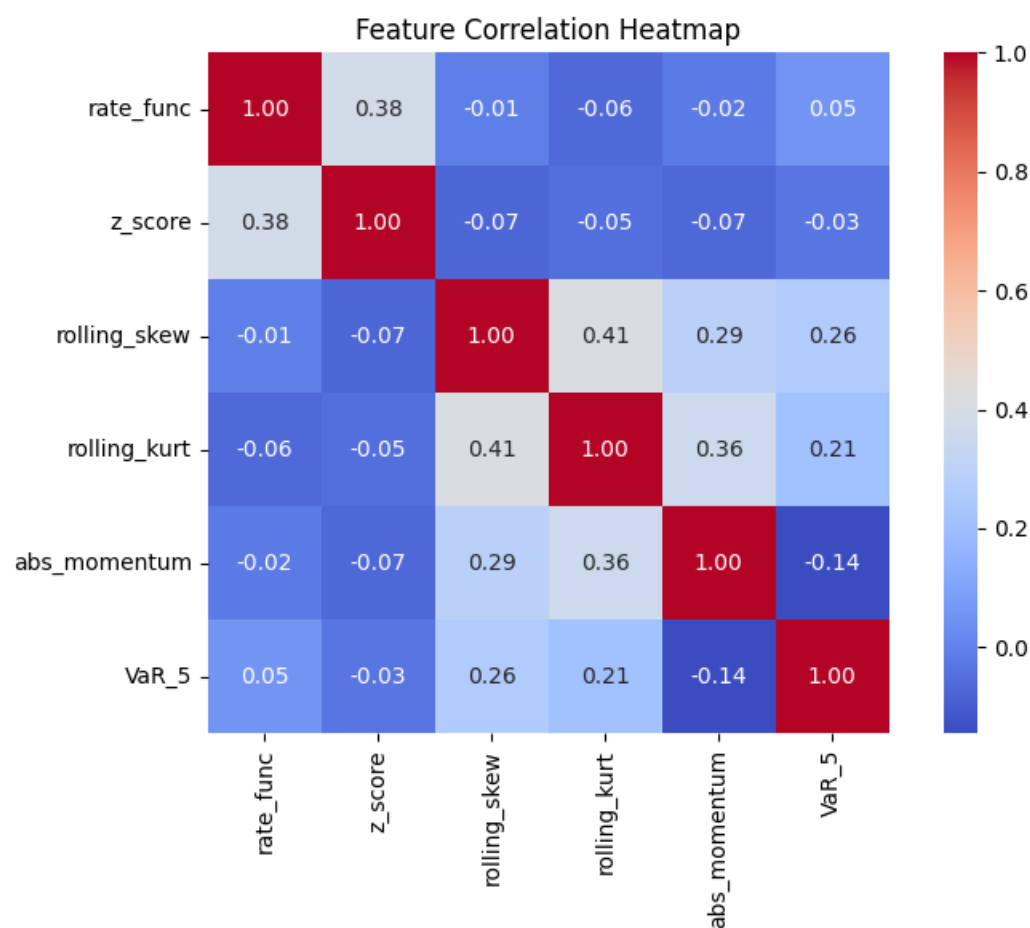
Logit Regression Results						
Dep. Variable:	extreme_event	No. Observations:	2468			
Model:	Logit	Df Residuals:	2462			
Method:	MLE	Df Model:	5			
Date:	Mon, 09 Jun 2025	Pseudo R-squ.:	0.2560			
Time:	06:07:07	Log-Likelihood:	-402.15			
converged:	True	LL-Null:	-540.49			
Covariance Type:	nonrobust	LLR p-value:	1.029e-57			
	coef	std err	z	P> z	[0.025	0.975]
rate_func	0.2164	0.031	6.922	0.000	0.155	0.278
z_score	-1.9835	0.131	-15.126	0.000	-2.240	-1.726
rolling_skew	-0.1617	0.049	-3.315	0.001	-0.257	-0.066
rolling_kurt	-0.0392	0.009	-4.179	0.000	-0.058	-0.021
abs_momentum	-203.6571	68.610	-2.968	0.003	-338.129	-69.185
VaR_5	90.5498	6.521	13.886	0.000	77.769	103.331

6.2 共線性與變數獨立性分析

為確保模型穩定性與指標解釋性，本研究進一步檢驗各風險指標間的相關性與共線性狀況。從圖二的皮爾森相關係數矩陣與圖三的變異數膨脹因子（VIF）檢定結果可見，所有指標間皆無明顯高度線性相關，亦未出現共線性風險。

這代表各項風險因子在模型中具有獨立貢獻性，彼此間的衡量邏輯不重疊，使模型能夠清楚分辨來自不同面向的潛在風險訊號，如報酬偏離、尾部厚度、分布傾斜性與歷史極端損失等，形成完整的多角度風險識別架構。

圖二：各指標共變異數矩陣



圖三：VIF(變異膨脹因子)觀察

VIF 結果：		
	Feature	VIF
0	const	5.980186
1	rate_func	1.181814
2	z_score	1.181708
3	rolling_skew	1.321346
4	rolling_kurt	1.372431
5	abs_momentum	1.296589
6	VaR_5	1.193991

6.3 分類表現與預測結果分析

本研究將邏輯回歸應用於每日市場風險預測，模型根據六項統計指標輸出每一日「發生極端事件」的機率分數，並以 0.5 作為分類閾值進行事件判斷。以下從整體分類績效、實際預測結果與時間分布三方面進行說明。

（一）分類指標表現

從圖四可見，模型在測試集（共 1310 筆樣本）上的分類績效如下：

- 整體分類準確率（accuracy）為 97.25%，表示大多數日期能被正確歸類。
- 針對極端事件（class 1）的 Precision 達 93.1%，Recall 為 62.8%，F1-score 為 0.75，顯示模型不僅能有效辨識過半數的極端事件，且誤報風險極低。
- Macro average 與 weighted average 的 F1 分數皆達 0.86~0.97，代表在類別不平衡（極端事件為少數）情況下，模型仍具穩定表現。

此結果凸顯本模型作為風險預測工具的兩大優勢：**高辨識度與低誤報率**，皆有助於實務上的風險因應。

（二）分類結果與錯誤分布

進一步從圖五混淆矩陣分析模型預測表現：

- 在 1224 筆正常日中，僅錯判 4 筆為極端日，準確率高達 99.7%。
- 極端事件共 86 筆，其中有 54 筆被正確辨識（Recall = 62.8%），漏判 32 筆，大多數接近臨界值。
- 誤判情形主要發生在跌幅接近 -5% 附近，顯示模型對邊界事件的判斷略有不穩，仍具未來精進空間。

整體而言，模型對於「安全時機」的辨識幾乎無誤，對「潛在風險日」的判斷則能抓住大多數關鍵日，具備有效實務應用潛力。

(三) 時間序列預測結果

圖五為模型每日預測風險分數（藍線）與實際極端事件（紅點）之時間序列對照圖。可明顯觀察到：

- 多數紅點發生前，藍色預測線有顯著上升趨勢，風險分數提前跳升具預警價值。
- 少數未能辨識的事件多為跳空型或無前兆事件，屬統計模型難以掌握的類型。
- 整體來看，風險分數變化與實際事件分布之間呈高度一致性，模型不僅能即時反應市場風險，亦具備「連續性輸出」特性，便於投資人設定彈性風險門檻。

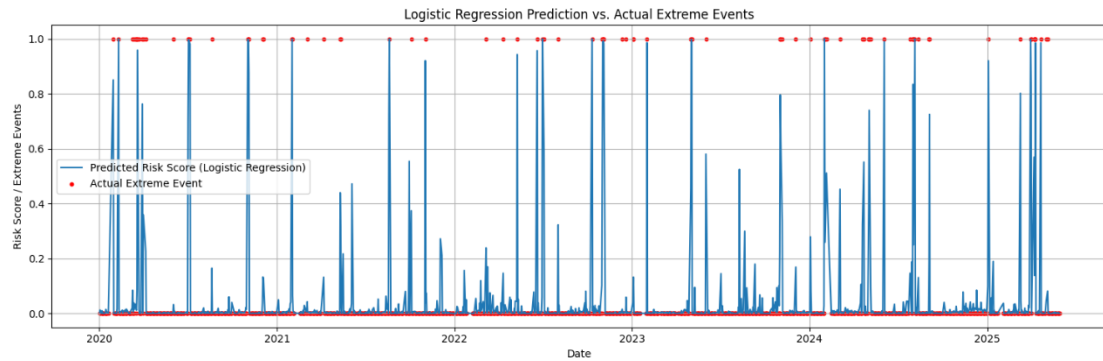
圖四：Logistic Regression 模型在測試集上指標之表現

	precision	recall	f1-score	support
0	0.974441	0.996732	0.985460	1224.000000
1	0.931034	0.627907	0.750000	86.000000
accuracy	0.972519	0.972519	0.972519	0.972519
macro avg	0.952738	0.812320	0.867730	1310.000000
weighted avg	0.971591	0.972519	0.970003	1310.000000

圖五：Logistic Regression 預測結果混淆矩陣

	預測:正常日	預測:極端日
實際:正常日	1220	4
實際:極端日	32	54

圖六：預測風險分數與實際極端事件之時間序列對照圖（2020 - 2025）



七、結論

本研究旨在建構一套具備前瞻性的市場風險預測機制，以協助投資人於極端下跌事件發生前，提早辨識潛在風險、啟動避險操作。研究過程中，綜合應用大離差理論（Large Deviation Principle）與統計風險指標（如 Z-score、Skewness、Kurtosis、VaR 等），並以邏輯回歸模型為基礎，進行每日極端事件機率之預測。

模型實證結果顯示：

- 所設計之六項風險指標皆達統計顯著水準，具備明確的經濟與風險解釋意涵；
- 在測試期間（2020 - 2025），模型可成功預測超過六成的極端下跌事件（Recall = 62.8%），且誤報率極低（Precision = 93.1%），整體準確率達 97.25%，分類表現穩定；

預測分數與實際市場風險事件之時間分布高度一致，展現出良好的預警能力與應用潛力。此外，本研究亦透過相關性分析與 VIF 檢定確認模型無顯著共線性問題，強化預測邏輯的穩定性與可解釋性。

綜合而言，本研究所提出之風險預測框架不僅具備理論基礎與模型可行性，亦能有效對應實務市場波動情境，具備進一步應用於動態資產配置、選擇權策略調整、或主動避險觸發機制的潛力。

八、未來展望

本研究初步驗證以大離差理論結合統計指標與機器學習模型進行市場極端事件預測的可行性，並在分類準確率與前瞻預測力方面皆展現良好成果。然而，為進一步強化實務應用與策略整合，後續研究可從以下三個方向持續深化：

8.1 結合策略應用，導入選擇權動態部位調整邏輯

未來研究可將本模型所輸出的風險預測分數，直接作為選擇權策略的判斷觸發依據。例如結合 Covered Call、Protective Put 等基本部位，並根據模型信號進行動態加減倉調整：

- 當模型預測高機率出現下跌風險時，結合 Buy Put 進行保護
- 當模型判斷市場穩定或波動收斂，可透過 Short Call（如 Covered Call）穩定收取權利金
- 若模型偵測潛在大漲徵兆，則結合 Buy Call 把握上行行情

此機制可望補強傳統策略無法應對系統性風險的缺陷，提升整體報酬與風險控制效率。

8.2 擴展模型結構與指標來源

目前模型基於簡單的邏輯回歸結構，後續可考慮導入更具表現力的模型，如 XGBoost、LSTM 或多階段分類器，以捕捉更細緻的市場非線性反應。同時，亦可嘗試將大離差理論推展至 Heston 或跳躍擴散類模型，結合價格波動與隱含跳點風險進行指標設計，提升對「突發性崩跌」的預警力。

8.3 精煉風險事件定義與強化模型適應性

目前極端事件定義以單日報酬低於 -5% 為準，雖具簡潔性與辨識性，但未來可進一步探討動態門檻設定（如以波動調整後的 VaR 為基準），或引入多維極端風險判準（如波動率跳升、成交量異常等），使模型能更貼近實際市場異常行為。

整體而言，未來研究將朝向策略實踐化、多模型結合與風險判準精緻化三大方向發展，最終目標為建立一套可自動監控風險、動態調整部位的量化投資系統，實現真正融合風控與資產配置的智慧決策框架。