Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-025-05920-7 OPEN

More than ex-post fitting: log-periodic power law and its AI-based classification

Ganghyeok Lee[1,2], Minhyuk Jeong[1,2], Taeyoung Park[3]✉ & Kwangwon Ahn[1,2]✉

The log-periodic power law (LPPL) model can be used to model the formation and evolution of endogenous bubbles in asset prices to predict the critical time of bubble bursts; however, it has a limitation in that the parameter interpretation and reliability are contingent on the researchers' subjective judgments. We introduce an artificial intelligence-based framework to address the reliability issue of the LPPL model. Specifically, we use an AI classification model to produce a reliability score, which is then incorporated into a novel risk metric designed to provide a more reliable early warning indicator of potential financial crashes. To train classification models, we created over 13 million labeled LPPL parameter sets estimated for daily closing prices of large-cap stocks in the United States (covering 24 years). We consider a simple formula to reformulate the crash information provided by the LPPL outcomes to create our risk metric, distance-to-crash with AI (DTCAI). When interpreting the LPPL results through our proposed risk metric, greater confidence can be placed on the predicted crash time as the metric inherently incorporates reliability information. Finally, the economic significance of the DTCAI metric is subsequently demonstrated through its application in dynamic portfolio-rebalancing strategies. The results show that strategies guided by our DTCAI metric significantly outperform the benchmark buy-and-hold strategy, regardless of the investor's risk preference. This study suggests that integrating artificial intelligence into financial models like LPPL can enhance the accuracy of bubble detection and risk quantification, offering a more robust approach to anticipating financial crises and optimizing investment strategies.

[1] Department of Industrial Engineering, Yonsei University, Seoul, South Korea. [2] Center for Finance and Technology, Yonsei University, Seoul, South Korea. [3] Department of Applied Statistics, Yonsei University, Seoul, South Korea. ✉email: tpark@yonsei.ac.kr; k.ahn@yonsei.ac.kr

Introduction

Detecting and predicting asset price bubbles and their eventual crashes have long challenged economists and financial practitioners. Financial bubbles are characterized by the rapid and unsustainable rise in asset prices beyond their fundamental values, followed by sudden and often catastrophic crashes, and they have had far-reaching consequences. From the Dutch Tulip Mania of the 17th century to the more recent 2008 Global Financial Crisis, it is critical to understand how bubbles form and burst to mitigate their impact on economies. The log-periodic power law (LPPL) is one mathematical model developed to address this issue. It has been used to predict the timing of the endogenous bubble crash based on precursor phenomena: (i) faster than exponential growth and (ii) increasing frequency of price oscillation (Geraskin and Fantazzini, 2013). The LPPL model is based on the theory that rational and irrational investors' herding behavior forms bubbles (Johansen et al., 2000), suggesting that speculative dynamics create an unsustainable growth pattern with increasing price volatility as the market approaches a crash, leading to a sharp reversal. Many prior studies have effectively predicted past financial market crashes using the LPPL model (Zhang et al., 2016; Dai et al., 2018; Ahn et al., 2024a, 2024b); however, despite its theoretical and empirical success, the LPPL model has limitations.

One main challenge of the LPPL model stems from its subjectivity and difficulty in parameter estimation and interpretation. The model requires the estimation of seven parameters from often noisy data, and minor variations in the input data (especially the size of the event window used for estimation) can lead to significant differences in the predicted outcomes (Feigenbaum, 2001). Furthermore, some log-periodic precursors do not necessarily lead to crashes (Laloux et al., 1999), leading to potential false signals. A significant issue is that the LPPL fitting procedure always yields a predicted critical time (t_c) for any given price series, regardless of whether a genuine bubble exists. Moreover, optimizing its parameters is a difficult task because the LPPL model is highly nonlinear. Hence, fitted LPPL models may indicate elusive crash dates, and it is crucial to rigorously assess the outcomes of the LPPL models before utilizing them in making financial decisions.

Researchers have taken various measures to overcome these problems. For example, Filimonov and Sornette (2013) presented a simple transformation of the LPPL model's formulation to reduce the parameters to three nonlinear parameters, thus decreasing the complexity of the fitting process. Moreover, Sornette et al. (2015) and Zhang et al. (2016) introduced DS LPPLS confidence and DS LPPLS trust indicators, which implement the bootstrap method to measure how well the LPPL model fits the price data. They also utilized shrinking time windows to average the volatility in parameter estimation and limited their scope of prediction to a change in regime (end of a bubble) rather than a crash. This approach allowed them to include cases where the asset price exhibits gradual drawdowns or even draws; however, their proposed indicator requires 12,500 repeated estimations of LPPL parameters and relies on experience-based filtering conditions to distinguish the good from the bad. Finally, Dai et al. (2018) employed diagnostic tests like the unit root test of residuals, sensitivity analysis of the parameters, and crash lock-in plot analysis to show the model's stability; however, their conclusions remain highly subjective to researchers' interpretations, and the tests were conducted only for cases where the model accurately predicted the crash date. They did not provide analysis for cases in which the model failed to predict the crash date; thus, the problem regarding the reliability of the LPPL's estimated parameters remains unsolved.

This study addresses this reliability problem by proposing a framework that integrates an artificial intelligence (AI) algorithm with the LPPL model. Specifically, we first fit the LPPL model to the log-price and obtain an LPPL parameter set. We then evaluate the reliability of the estimated LPPL parameter set, particularly the critical time, using an AI-based classification approach. The fundamental assumption underpinning this approach is that the specific combination across the seven estimated LPPL parameters holds discernible patterns reflecting the quality of the fit. We posit that LPPL aligns with genuine market bubbles, and the imminent crashes yield parameter sets with different quantitative characteristics compared to parameter sets resulting from spurious fits or fits obtained during non-bubble market phases. The AI model is essentially trained to recognize these complex patterns. To this end, we first collected a comprehensive dataset of over 13 million labeled LPPL parameter sets. Then, we trained and compared popular classification models, such as the artificial neural network (ANN), random forest (RF), and logistic regression. While training the classification models, we used the LPPL parameter sets as input and the labels of these LPPL parameter sets, which take 1 if the predicted crash date is within 10 days of the actual crash date and 0 otherwise, as output. Once the classification models are trained, they can calculate the reliability score of any LPPL parameter set objectively and with low computational resources and time. Next, we propose a risk metric that measures the distance to the nearest price crash. The risk metric is then multiplied by the reliability score obtained from our trained classification model to prevent false crash alerts from unreliable LPPL model outputs. Accordingly, we expect that integrating LPPL and AI may help reduce false alarms generated by LPPL predictions. Finally, we verify and test a possible avenue for utilizing our proposed risk metric in a simple stock/bond portfolio-rebalancing strategy. Specifically, we construct three dynamic portfolio-rebalancing strategies based on an investor's risk preference; we incorporate the risk metric as the rebalancing signal and compare the results to static buy-and-hold benchmarks. The results show that our strategy outperforms the benchmark regardless of the investor's risk preference, highlighting that our metric provides valuable information that can be used to hedge against rare but critical crash risk.

The contribution of this research is threefold. First, we incorporate AI into the LPPL framework, addressing a key limitation of the existing bubble prediction models and improving the parameter estimation's consistency and accuracy. We train three well-known classification models in finance and compare their performances in out-of-sample prediction. Second, we propose a novel risk indicator based on the results from the LPPL model estimation and the best-performing classification model in the previous step. This indicator measures the risk of asset price crashes, quantifying the distance to a significant drawdown or crash. Third, we test this risk indicator within a portfolio-rebalancing strategy. The results demonstrate that the portfolios using our risk metric are superior regarding the Sharpe ratio and returns compared to the benchmark portfolios.

The rest of the paper is organized as follows: "Literature review" thoroughly reviews the related studies, "Models" presents the LPPL model and the classification algorithms, "Distance-to-crash: A new risk metric" introduces our risk metric. "Empirical analysis" provides the data, explains our methodological framework, and presents the portfolio-rebalancing experiment design and results. Finally, "Conclusion" concludes.

Literature review

Financial risk. All financial assets except risk-free bonds are exposed to systematic, volatility, tail, and jump risks. Financial risks are measured by quantitative risk metrics to monitor and

manage the potential loss or adverse outcomes of an investment or financial activity, allowing firms and investors to allocate capital efficiently. A "good" risk metric should be robust, intuitive, and capable of capturing various potential outcomes under different market conditions; it should also be forward-looking, computationally feasible, and adaptable to various asset classes.

Systematic risk originates from the risk inherent in the entire market or macroeconomy, usually measured through factor-based linear models. For example, the market beta is a measure of systematic risk that quantifies the sensitivity of an asset's or portfolio's excess returns to market movements. It is a key component in the capital asset pricing model (CAPM), which connects an asset's expected return to its market risk exposure (Sharpe, 1964); however, this measure only captures a single dimension of risk, namely, market risk, and does not consider other risk factors significantly affecting asset returns. Previous studies extended the CAPM to address these limitations, expanding the analysis beyond market risk to include additional factors to explain asset returns. One CAPM extension is the Fama–French three-factor model, which incorporates size and value factors, recognizing that small-cap stocks and undervalued stocks tend to exhibit distinct risk-return characteristics (Fama and French, 1996). Other factor-based risk models have added momentum, profitability, and investment factors, among others, to capture a broader range of risks (Ehsani and Linnainmaa, 2022; Fama and French, 2015). These multifactor approaches provide a more nuanced understanding of various systematic risks; however, systematic risks cannot be eliminated through diversification. Therefore, the purpose of measuring systematic risks is usually to price an asset or portfolio rather than to quantify the upfront financial risk.

Some studies have suggested mathematical models for financial risks. One basic approach is modeling an asset price process as a random walk (RW). For example, a geometric Brownian motion is the continuous-time analog of a log-normal random walk. Its logarithm follows a Brownian motion with drift, which is a continuous-time version of a random walk with drift. The log returns are normally distributed; thus, we can measure the volatility risk as the diffusion coefficient in the GBM model, which can be calibrated as the standard deviation (SD) of log returns or the realized volatility based on high-frequency data (Andersen et al., 2003). A higher SD indicates greater volatility and, thus, increased risk associated with the asset. Furthermore, we can easily measure tail risks like value-at-risk (VaR) and expected shortfall (ES) based on the RW model; under the RW model, VaR is estimated as a tail quantile, and ES is estimated as the conditional expectation given a VaR violation in a normal distribution. Conversely, in financial markets, volatility clusters and significant drawdowns occur more frequently than equally large upward movements (Cont, 2001), resulting in non-zero skewness and excess kurtosis of asset return distributions. Empirical evidence has revealed that financial asset returns are not normally distributed, e.g., in the stock (Cont, 2001) and crude oil futures markets (Jeong et al., 2023); therefore, measuring financial assets' risks based on the RW model cannot reflect these stylized facts.

Time-varying volatility models have been developed to capture the time variation of return volatility, where the diffusion coefficient is assumed to be time-varying, unlike the RW model. For example, conditional volatility models, such as autoregressive conditional heteroskedasticity (ARCH) (Engle, 1982) and generalized ARCH (Engle and Bollerslev, 1986), can capture features like volatility clustering and persistence. More advanced methods, such as the stochastic volatility model (Heston, 1993), offer additional insights into the nature of volatility risk; however, in financial markets, we occasionally observe sudden, discontinuous changes in asset prices. Such discrete jumps can be modeled by considering a jump process for asset prices. For example, Merton (1976) proposed a jump-diffusion model that adds a jump process to the diffusion process model. Eraker et al. (2003) further showed that the model incorporating jumps into both price and volatility processes could explain the rapid changes in volatility risks. When measuring financial risks, we should determine which model to describe the price and volatility processes for the asset of interest according to its market characteristics and the modeling purpose. Once the processes for an asset's price and volatility are modeled, we can compute the volatility, tail, and jump risks that the asset is exposed to through simple calculations or simulations.

The LPPL model (Johansen et al., 2000) starts with a jump process for asset prices to model their downside jump risks, i.e., crash risks. In particular, the LPPL model offers insights into the timing of a potential market crash; conversely, most traditional measures of financial risks primarily quantify the magnitude of loss or potential fluctuations and the probability of extreme events. We use the predicted time of a market crash from the LPPL model and introduce a new financial risk metric that captures the temporal aspect of risk, aiming to measure an asset's or portfolio's proximity to a sharp downturn. By measuring the time horizon to a potential crash, we provide more actionable information for risk management, allowing for timely adjustments to portfolio strategies to mitigate catastrophic losses. This approach can help investors and risk managers who seek to anticipate market turning points rather than understand the potential magnitude of their risks.

This study aims to propose a risk metric that is an early warning indicator, signaling potential vulnerability to large drawdowns before they occur. Existing early warning indicators have focused on potential crises in the currency or banking system (Aldasoro et al., 2018; Bussiere and Fratzscher, 2006; Edison, 2003; Ji et al., 2020). In contrast, the risk metric proposed in this study concentrates on the timing of crash risk in an individual asset's or portfolio's value using the LPPL model. Previous studies suggested that the LPPL model provides early warning signals of financial crises. For example, Kurz-Kim (2012) reported that early warning signals could be captured by observing the magnitude and the amplitude decomposed from the LPPL model. Zhang et al. (2016) proposed several innovations in using LPPL (e.g., introducing quantile regression and bubble diagnostics) and suggested that those innovations can help the LPPL model be included in an early warning system. Nonetheless, these studies do not propose an explicit early warning system based on the LPPL model, primarily due to the substantial computational complexity and the considerable time required to robustly estimate and validate a large number of LPPL fits for real-time application. Conversely, we propose a forward-looking risk metric that captures whether a bubble exists in an asset's or portfolio's price and how close that bubble is to the crash. Our risk metric is an effective early warning indicator that provides signals that directly convey an intuitive and understandable measure of crash proximity. Furthermore, although the initial data generation and AI model training stages in our framework are computationally intensive, the subsequent daily computation of the risk metric is highly efficient and fast, making the framework suitable for real-time market monitoring.

Classification models in finance. Classification models have become integral to finance, where they are used to predict categorical outcomes like credit risk, stock market movements, and bankruptcy probability. By assigning financial entities to specific categories, classification models assist in decision-making

processes essential for risk management, portfolio optimization, fraud detection, and credit scoring. ANN, RF, and logistic regression are among the most prominent models in finance, each offering distinct advantages.

ANNs have gained considerable attention in finance because they can capture nonlinear relationships in complex datasets through multiple layers of interconnected nodes (LeCun et al., 2015). Several studies have highlighted ANN's advantages over other models in predicting financial crises, stock market trends, and credit risk. For example, Atiya (2001) used ANN to predict bankruptcy, finding that it performed better than statistical models like logistic regression. Singh and Srivastava (2016) applied deep neural networks to model stock market movements. They found that ANN's superior ability to capture non-linearities yielded higher predictive accuracy than traditional models.

RF, on the other hand, is an ensemble learning method that utilizes multiple decision trees and aggregates their prediction to improve classification accuracy (Breiman, 2001). Additionally, Lessmann et al. (2015) showed that RF outperforms traditional models in predicting credit risk, default probability, and stock market classifications. Finally, Xu et al. (2021) demonstrated RF's effectiveness in predicting loan defaults, highlighting its capacity to capture complex relationships that simpler models miss.

Logistic regression has long been one of the most widely used classification models in finance due to its simplicity, interpretability, and efficiency. For instance, Ohlson (1980) applied logistic regression in a pioneering study of bankruptcy prediction, while Chen (2011) utilized logistic regression to predict corporate financial distress. Finally, Dutta et al. (2012) used logistic regression to determine the indicators that significantly affect the performance of companies in the stock market.

Considering their distinctive strengths and widespread use in financial studies, we selected the three classification models above and trained them to evaluate the reliability of the estimated LPPL models. ANN was chosen for its ability to capture complex nonlinear relationships and interactions between variables. RF was included for its robustness and effectiveness in reducing overfitting through bootstrapping and feature randomness. Finally, logistic regression was selected for its simplicity and efficiency in binary classification tasks. By comparing these three models, we have encompassed the most widely recognized classification techniques currently available.[1]

Models

Log-periodic power law. The LPPL model is frequently employed to detect financial bubbles and predict their associated crashes (Dai et al., 2018). In this model, traders form groups exhibiting self-similar behavior through interactions (Geraskin and Fantazzini, 2013; Jang et al., 2018a). Specifically, traders' buy and sell decisions are influenced by the actions of others within their neighborhood. Some traders make overconfident investment decisions, which others in the same network imitate, leading to a boom cycle (Lee et al., 2020). Rising asset values cause speculative reinvestment, and this self-reinforcing loop repeats over time (Dai et al., 2018; Jang et al., 2020) until a certain point, called the critical time; the LPPL model includes this time as one of its parameters (Zhou and Sornette, 2006; Jang et al., 2018b, 2020). In the following, we introduce the mathematical derivation of the LPPL model.

Price dynamics. The LPPL model starts with the following jump process for an asset's price:

$$dp_t = \mu_t p_t dt - \kappa p_t dj, \tag{1}$$

where p_t is the asset price, μ_t is the drift term, dj represents a discontinuous jump, and κ is the jump size. j is a dummy variable that takes one after the critical time and zero otherwise. Assuming the rational expectations and the no-arbitrage condition, the process in Eq. (1) would be martingale, such that

$$E_t\left[dp_t\right] = \mu_t p_t dt - \kappa p_t h_t dt = 0, \tag{2}$$

where $h_t \equiv E_t\left[dj/dt\right]$ represents the crash hazard rate. Accordingly, we have $\mu_t = \kappa h_t$, which implies that rational traders accept the risk of a bubble crash if the price increase is high enough to compensate for the risk. Before the crash, $dj = 0$, then the stochastic differential equation in Eq. (1) can be rewritten as an ordinary differential equation, $d\ln p_t = \kappa h_t dt$, whose solution is as follows:

$$\ln\left(p_t/p_{t_0}\right) = \kappa \int_{t_0}^{t} h_u du. \tag{3}$$

Modeling the crash. Johansen et al. (2000) assumed that a crash occurs because traders imitate their nearest neighbors. Moreover, each trader i's state, s_i, can only be either $+1$ or -1, which is determined according to

$$s_i = \mathrm{sign}\left(K \sum_{j \in N(i)} s_j + \sigma \varepsilon_i + G\right), \tag{4}$$

where K and σ are the tendency toward imitation and idiosyncratic behavior, respectively. $N(i)$ is the set of trader i's neighbors that interact with the trader i, and ε_i represents the trader i's idiosyncratic behavior following the standard normal distribution. G represents the global influence on traders' states. Johansen et al. (2000) defined the system susceptibility as follows:

$$\chi = \left|\frac{d(E[M])}{dG}\right|_{G=0}, \tag{5}$$

where $M \equiv \left(1/I\right)\sum_{i=1}^{I} s_i$, with I being the number of all traders in the system. Then, they assumed a hierarchical diamond lattice for the traders' network structure and showed the first-order expansion of the general solution (Derrida et al., 1983):

$$\chi \approx a_1\left(K_c - K\right)^{-\gamma} + a_2\left(K_c - K\right)^{-\gamma}\cos\left[\omega \ln\left(K_c - K\right) + \psi\right], \tag{6}$$

where a_1, a_2, ω, and ψ are constant real numbers. γ is the critical exponent of the susceptibility, K represents the coupling strength between neighbors, and K_c is the critical point. Denoting the critical time of the crash as t_c, Johansen et al. (2000) approximated $t_c - t$ to a multiple of $K_c - K$ before the crash (i.e., $t < t_c$). Based on this approximation, they assumed that the behavior of the crash hazard rate, h_t, has the same mathematical form as that of susceptibility. This assumption resulted in the following approximation of h_t:

$$h(t) \approx b_1\left(t_c - t\right)^{-\alpha} + b_2\left(t_c - t\right)^{-\alpha}\cos\left[\omega \ln\left(t_c - t\right) + \varphi\right], \tag{7}$$

where b_1, b_2, ω, and φ are constant real numbers, and α is the exponent of the power law growth.

The solution to the LPPL model. Using Eqs. (3) and (7), we can obtain the solution to the LPPL model in the following form (Johansen et al., 2000):

$$y_t = A + B\left(t_c - t\right)^{\beta}\left\{1 + C\cos\left[\omega \ln\left(t_c - t\right) - \phi\right]\right\} + \epsilon_t \text{ for } t < t_c, \tag{8}$$

where $y_t > 0$ is the log price at time t. $A > 0$ is the log price at the critical time t_c, and $B < 0$ is the increase in y_t over the time before the crash when C is close to 0. $C \in [-1, 1]$ controls the magnitude of the oscillations around the exponential trend, and $t_c > 0$ is the critical time. $\beta \in [0, 1]$ is the exponent of the power law growth. $\omega > 0$ is the frequency of the fluctuations during the

Table 1 LPPL parameters and their descriptions.

Parameter	Constraint	Description
A	$(\max P, +\infty)$	The theoretical price at the critical time, which must be greater than the maximum of the time series, P used for LPPL fitting
B	$(-\infty, 0)$	Negative, measuring the distance to the critical price
C	$C \in (-1, 1)$	Oscillations cannot overwhelm the trend
t_c	$(t, +\infty)$	(1) The most probable critical time; (2) Have not crashed
ϕ	$[0, (2\pi)$	Phase adjustment
ω	[4.8, 13]	The frequency of the fluctuations during the bubble Too small: False oscillation Too large: Fit the random noises
β	[0.1, 0.9]	The exponent of the power law growth Negative β: Infinite critical price Positive but close to 0: No trend Positive but close to 1: Low hazard rate

bubble, $\phi \in [0, 2\pi]$ is a phase parameter, and ϵ_t is Gaussian white noise. Table 1 summarizes the LPPL parameters, descriptions, and the constraints imposed on those parameters while fitting the LPPL models to historical observations.

We selected seven parameters that minimize the root mean square error (RMSE) between the observed and predicted values of the LPPL model:

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t - y_t\right)^2}, \tag{9}$$

where Y_t is the model response at time t, and T is the number of observations.

This study produced initial values for the parameters using the price gyration method and optimized them using a genetic algorithm (GA) (Dai et al., 2018; Jang et al., 2018a). The price gyration method generates the initial parameters by first identifying three consecutive peaks, i, j and k, in the time series over varying window sizes. We use a distance-based weighted approach to prioritize the more recent peaks.[2] Then, the initial values of t_c, ω and ϕ are calculated as $t_c = (pk - j)/(\rho - 1)$, $\omega = 2\pi/\ln\rho$, and $\phi = \pi - \omega\ln(t_c - k)$ with $\rho = (j - i)/(k - j)$. The initial values of β and C are set to 1 and 0, respectively. The remaining parameters, A and B are estimated using an ordinary least squares fit:

$$y_t = A + B\left(t_c - t\right) + \epsilon_t, \tag{10}$$

where ϵ_t is the error term. A more detailed explanation can be found in Dai et al. (2018).

GA is a nonlinear optimization algorithm inspired by natural selection and evolution biology principles. Starting from an initial population, potential solutions, called individuals, evolve over successive generations through specific operations. These operations include selection (choosing the fittest individuals), crossover (recombining parts of two individuals to form new ones, usually by taking the mean), and mutation (randomly adding deviations to parts of an individual to introduce variation). Each individual's fitness was evaluated based on a predefined objective function, the RMSE in this study.

We take the following steps to generate the LPPL parameter sets, which will be used to train the classification models. First, given a log price series, the price gyration method generates an initial population of 200 LPPL parameter sets. Second, starting from this initial population, the fitness score (measured in RMSE) of all 200 LPPL parameter sets (hereafter referred to as individuals) is calculated. We then search for the optimal LPPL parameter set using the GA:

(a) The top 5% of individuals (based on RMSE) are passed directly to the next generation as elite children, ensuring that the best solutions found so far are preserved.
(b) From the population, the parents are selected by the stochastic universal sampling method (Baker, 1987). This method lays out a line in which each parent corresponds to a section of the line of length proportional to its fitness score; the algorithm moves along the line in steps of equal size. At each step, the algorithm allocates a parent from the section it lands on, allowing fitter individuals to have a higher probability of being selected. The total number of parents selected was set to $2n_c + n_m$, where n_c is the number of crossover children and n_m is the number of mutation children.
(c) Crossover children are generated by taking the arithmetic average of the parameter vectors of randomly selected pairs of parents. Mutation children are created by applying random changes to individual parents; specifically, we add a random vector drawn from a Gaussian distribution to the parent's parameter vector. The ratio between the number of crossover and mutation children was set to 8:2. In summary, 10 elite individuals, 152 crossover children, and 38 mutation children were passed on to the next generation.
(d) The process through (a)–(c) is repeated for 700 generations.

Finally, we select the parameter set with the lowest RMSE among the individuals in the population of the 700th generation. This approach results in one suboptimal or potentially optimal LPPL parameter set for the given log price series.

To improve the classification model training, we need a large dataset. Therefore, for one log price series, we repeat the above parameter estimation framework, from the initial population generation to the 700-generation GA, 500 times to produce 500 sets of LPPL parameters fitted to the given log price series. Although these 500 parameter sets are estimated from the same log price series, they are sufficiently distributed to the extent that they can provide classification models with different information because the GA is a heuristic algorithm whose final output can vary according to hyperparameters and the initial population.

Classification models. The classification models assess the reliability of a given set of LPPL parameters, particularly concerning the predicted critical time, t_c. The models are trained to learn the complex patterns within the seven LPPL parameters (estimated via the described gyration and GA process) that distinguish between LPPL fits historically associated with an actual crash (our "reliable" class, label 1) and those that were not ("unreliable" class, label 0). Thus, the classification aims to produce a "reliability score" for the estimated t_c from an input of LPPL parameter set. To achieve this, we trained and tested three widely used classification models (ANN, RF, and logistic regression) and compared their performance in classifying the reliability label based on the input parameter set.

Mathematically, we can express this classification framework using a functional form:

$$P = f(X; \Theta), \tag{11}$$

where X is the 7×1 vector of the LPPL parameters. $P \in [0, 1]$ is the reliability score (predicted output), f is the classification model used, and Θ represents the parameters of the classification model (f). The form of f and Θ depends on the model used: ANN, RF, or logistic regression. The objective of training and testing these models is to select Θ that yields the best predictive performance, evaluated using the recall metric. In the context of

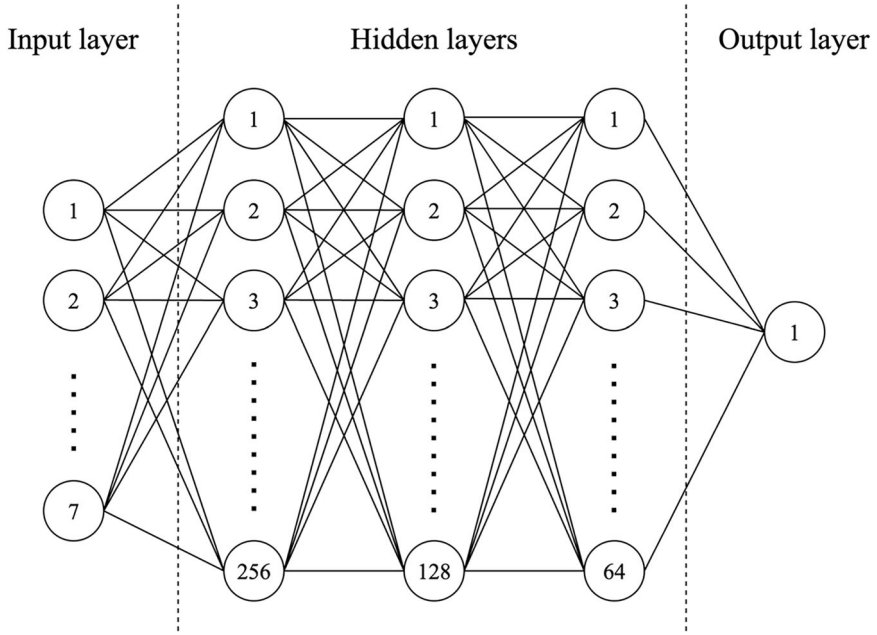

Fig. 1 Architecture of the ANN. The circles represent nodes, and the lines represent the connections (weights) between them.

financial crash prediction, the consequences of failing to predict a crash (a Type II error or false negative) are far more severe than the costs associated with a false alarm. Missing a true crash signal can lead to significant, unprepared financial losses for investors and institutions; therefore, maximizing recall—the proportion of actual crash events the model correctly identifies—is paramount. Other metrics exist and are presented in the results for robustness; however, prioritizing recall ensures that the model is optimized to capture as many true warning signs as possible.

Artificial neural network. ANN comprises interconnected layers of nodes, known as neurons, where each neuron processes the input and produces an output based on an activation function. The ANN architecture includes an input layer, one or more hidden layers, and an output layer. Adjusting the weights of the connections between neurons through backpropagation enables ANNs to recognize patterns, classify data, and make predictions. In this study, the input layer comprises the estimated LPPL parameters; the output layer is a single value between 0 and 1, signifying the reliability score of these parameters. The rectified linear unit activation function is applied after each hidden layer, introducing nonlinearity; the sigmoid function is used in the output layer to ensure that the output value falls within the range of 0 to 1. The Adam optimizer combines the benefits of adaptive learning rates and momentum to improve convergence speed and stability and is used for backpropagation during model training. Binary cross-entropy is used as the loss function because it is well-suited for binary classification tasks, and recall is used as the primary evaluation metric to minimize the occurrence of false negatives.

The model's hyperparameters are optimized using the grid search method, which explores a predefined set of hyperparameter values by systematically and exhaustively testing each possible combination. This study used the recall on the validation set as the evaluation metric. The number of layers, size of each layer, batch size, and learning rate were optimized using the grid and had the optimal values of 3, {256, 128, 64}, 512, and 0.01, respectively. Figure 1 illustrates the finally selected architecture of the ANN.

Random forest. RF is an ensemble learning method first proposed by Ho (1995). It is used for classification, regression, and other tasks that operate by constructing multiple decision trees during training and outputting the mode of the classes or the mean prediction of the individual trees. The model improves performance by reducing overfitting (a common issue in single decision trees) by creating numerous trees that each vote on the outcome. RF leverages the "bagging" technique, where data subsets are randomly sampled with replacement to train individual trees, and a random subset of features is selected for splitting at each node. This randomness makes the RF robust and less sensitive to noisy data, offering higher accuracy and generalization than traditional decision trees (Breiman, 2001).[3]

The primary hyperparameters of RF are the number of decision trees, the maximum depth of the tree (max-depth for short), the minimum number of samples required to split an internal node (minsamples-split for short), the minimum number of samples needed to be at a leaf node (min-samples-leaf for short), and the number of features to consider when looking for the best split (max-features for short). This study set the number of decision trees to 500, following Breiman (2001). Like the ANN's hyperparameter tuning method, a grid search determines the optimal hyperparameters. The model's performance, measured using recall, is assessed for each combination in the hyperparameter space. After many experiments, this paper sets the max-depth, min-sample split, min-samples-leaf, and max-features as 20, 10, 10, and 40.

Logistic regression. Logistic regression is a widely used statistical method for binary classification tasks, where the goal is predicting the probability that a given input belongs to one of two classes. Unlike linear regression (which predicts continuous values), logistic regression applies a logistic function (sigmoid function) to model the relationship between the input features and the binary outcome. The sigmoid function maps predicted values to probabilities between 0 and 1, allowing for the classification of outputs based on a defined threshold (typically 0.5). The model can be mathematically expressed as follows:

$$P = f_{LR}(X;\alpha) = \frac{1}{1+\exp(-\alpha' X)}, \tag{12}$$

where α is the 7×1 vector of the parameter weights learned during model training; and P is the predicted probability,

representing the reliability score in our context. Despite its simplicity, the logistic regression performed well with linearly separable data.

Distance-to-crash: a new risk metric

We propose the following risk metric, distance-to-crash (DTC), to quantify the proximity of an asset price crash. DTC is the ratio of the time elapsed within a subseries to the time to the predicted crash, expressed as follows:

$$DTC = \frac{t_2 - t_1}{t_c - t_1}, \tag{13}$$

where t_1 represents the start of the subseries, t_2 is the end time, and t_c is the critical time of the crash, as predicted by LPPL. This metric captures the relative distance from the current time (t_2) to the anticipated crash (t_c), normalized by the total duration of the subseries. Thus, the DTC value's range is fixed between 0 and 1. Values near 1 indicate that t_2 is close to t_c, signaling a high risk of an imminent crash. Conversely, values closer to 0 imply a lower risk, as t_2 is further from the predicted crash time, t_c. DTC provides a fixed-bound metric, allowing a standardized method of using the estimated parameters of the LPPL model.

We adjust DTC using the reliability score (P) from the trained classification models to improve the reliability of our proposed risk metric, which we call DTC with AI (DTCAI). This adjusted metric multiplies the original DTC score by the output of our classification model, providing a reliability score between 0 and 1 based on the reliability of the predicted crash time, t_c. The metric is expressed simply as follows:

$$DTCAI = DTC * P. \tag{14}$$

We chose this multiplicative form, $DTCAI = DTC * P$, primarily due to its simplicity and intuitive interpretation. This approach directly scales the raw DTC metric (derived from the LPPL's predicted t_c) by the AI-assessed reliability score (P) of that prediction. When the AI model deems the LPPL fit reliable (P approaches 1), the DTCAI score remains close to the original DTC. Conversely, when the fit is unreliable (P approaches 0), the score is significantly reduced, effectively filtering or dampening the signals from less credible LPPL fits.

While we acknowledge the potential merit of alternative formulations for future research, we chose the $DTCAI = DTC * P$ formulation due to its parsimony, clear interpretability, and direct way of integrating the temporal risk information from DTC with the reliability assessment from the AI model (P).[4] This risk metric retains the full DTC value when the predicted crash time is deemed reliable by the classification model; however, it is scaled down when the reliability is lower. The adjustment effectively diminishes the scores of less reliable crash predictions, reducing the probability of false alarms. By combining the temporal information from the LPPL model with the reliability assessment of our classification model, DTCAI offers a more nuanced evaluation of crash risk. This approach ensures that high-risk signals are only issued when the timing of the crash is potentially imminent, and the reliability score of the prediction is high.[5]

Empirical analysis

Model training

Data. We first generated an extensive collection of estimated LPPL parameters (used as the selected models' input features) to train the classification models. Specifically, we fit the LPPL model using the daily closing stock prices of the top 100 companies among the S&P 500 index constituents regarding market capitalization as of September 19, 2023. This dataset was sourced from Datastream and spans 24 years, from January 1995 to December 2018. The sample period incorporated a range of market conditions, including periods of growth, recessions, and market turbulence. Notable events include the build-up and crash associated with the Dot-com in the late 1990s and the Global Financial Crisis (2007–2009). These major market cycles contained distinct bubble dynamics and crashes, providing critical historical examples for training the AI classification models to distinguish reliable patterns effectively.

LPPL model fit and preprocessing. The training of the classification model for evaluating the reliability of the estimated LPPL outcomes includes multiple steps. First, we use the log price data to fit the LPPL model following the estimation procedure described in "Log-periodic power law". We extract subseries for each stock price series using the sliding window technique with a window size of 504 days (2 trading years) and a step size of 21 days (1 trading month). For each subseries, we follow Dai et al. (2018) and use the price gyration and GA to generate 500 sets of LPPL parameters ($A, B, C, t_c, \phi, \omega$, and β) at each window. Following this method, we can generate over 13 million LPPL parameter sets, which we then use to train and test the classification models. To create the dataset for supervised learning, we applied a binary labeling function to generate a labeled dataset of LPPL parameters. The label is assigned a value of "1" (indicating a "reliable" fit in terms of crash prediction) if the predicted critical time (t_c) from that specific LPPL parameter set falls within 10 trading days of a historical crash date. If the predicted t_c does not meet this condition, the label is assigned a value of "0" ("unreliable"). An "actual historical crash date" is defined as the date of a price peak that satisfies two conditions based on Brée and Joseph (2013): (1) no closing price in the preceding 262 weekdays (approximately 1 year) exceeds the peak price, and (2) the stock price decreases by more than 25% within the 60 weekdays following the peak. These binary labels serve as the ground truth that the classification models aim to provide the crash signal based on the input LPPL parameter set.

Table 2 summarizes the labeling results, revealing a heavy class imbalance since crash-related ("reliable") labels are rare compared to non-crash ("unreliable") labels. This imbalance can bias classification models toward the majority class, leading to poor performance in identifying the crucial minority (crash) class (He and Garcia, 2009). To address this issue, we apply the synthetic minority oversampling technique (SMOTE) strictly to the training dataset (Chawla et al., 2002). SMOTE generates synthetic examples of the minority class (reliable LPPL parameter sets) by interpolating between existing real minority class instances. This approach provides the classification models with a more balanced view during the learning phase, improving their ability to recognize the characteristics of the rare but critical reliable/crash class.

We acknowledge the potential concern that oversampling techniques could introduce artificial patterns; however, it is crucial to note that SMOTE was applied only to the training data. The validation and test datasets used for model selection and final performance evaluation consist entirely of original, non-synthetic

Table 2 The number of LPPL parameter sets split by their label.

Label	Number of LPPL parameter sets
Unreliable (0)	13,423,859
Reliable (1)	426,123
Total	13,849,982

Table 3 Performance comparison of AI classification models for LPPL parameter reliability assessment on the test dataset.

Model	Accuracy	Recall	Precision	F1 score
ANN	55.41%	59.15%	42.41%	46.20%
Random forest	49.72%	1.58%	42.61%	3.05%
Logistic regression	50.71%	41.20%	50.88%	45.53%

data points. This approach ensures our reported performance metrics reflect the model's generalization ability to real-world data distributions. SMOTE generates synthetic feature vectors (LPPL parameters); however, it does not alter the underlying market events or the relationship between LPPL patterns and subsequent market outcomes in the real testing data. The reasonably good performance observed on the untouched test set and in the out-of-sample period ("Out-of-sample classification") suggests that the models learned meaningful patterns indicative of crash risk rather than artifacts solely attributable to the synthetic training examples.

Model performance. When evaluating the performance of the three classification models on the test set, we primarily focus on maximizing recall. As discussed in "Modeling the crash", minimizing Type II errors (failing to predict a crash when one occurs) is crucial given the potentially catastrophic financial consequences of missing such an event. Standard accuracy is not a suitable metric here due to the significant class imbalance present in our dataset (Table 2 shows far more "unreliable" labels than "reliable" ones). A classification model can achieve high accuracy simply by always predicting the majority class ("unreliable" or 0), rendering it useless for detecting rare but critical crash events (often referred to as the accuracy paradox in imbalanced datasets). Therefore, recall, which measures the model's ability to correctly identify the positive ("reliable" or crash) instances, is the most relevant criterion for evaluating model effectiveness in this specific crash-prediction context.

Table 3 reports the performance of the three classification models. The ANN model achieved a recall score of 59.15%, indicating a relatively strong ability to identify actual crashes compared to other models; however, this emphasis on recall sacrifices precision, leading to an F1 score of 46.20%, the harmonic mean of precision and recall. The ANN model outperforms the other two classification models in both recall and F1 scores, making it the most reliable choice for applying our DTCAI risk metric; therefore, the subsequent analyses were conducted using the trained ANN, the best-performing model.

Out-of-sample classification. We investigate the out-of-sample performance of DTC and DTCAI using the S&P 500 index data over six years (2019–2024). The LPPL parameters for the S&P 500 index are estimated using two-year rolling windows; the scores for both metrics are then calculated for each window. These scores represent the DTC and DTCAI of the last day in the corresponding window. These out-of-sample signals are generated using only data available up to the calculation date and the pre-trained AI model; this approach ensures no backward relabeling or look-ahead bias in this evaluation.

The calculated scores are categorized into three groups and labeled as follows. Scores above 0.6 are labeled "crash-alert," those between 0.3 and 0.6 are "caution," and scores below 0.3 are "stable." Table 4 summarizes the labeling results, indicating that DTCAI produces significantly lower scores overall than DTC, as expected, since DTCAI is the adjusted score that is reduced for unreliable model fits and preserved for reliable ones. DTCAI demonstrates a reduced proportion of crash alerts while the proportions of "caution" and "stable" increased. When evaluating the effectiveness of the DTC and DTCAI signals against actual market events in the extended period, we identify actual crashes using the exact definition employed for labeling our training data ("LPPL model fit and preprocessing"): the date of the price peak followed by a decline of more than 25% within the subsequent 60 weekdays (Brée and Joseph, 2013).

Table 4 Summary of the DTC and DTCAI score labels.

	Stable	Caution	Crash-alert
DTC	559 (38.0%)	552 (37.5%)	360 (24.5%)
DTCAI	725 (48.0%)	586 (38.8%)	199 (13.2%)

The numbers in parentheses indicate the proportion of the corresponding labels.

Figure 2 illustrates the time evolution of the DTC and DTCAI labels, presenting a clear picture of their differences. To quantify the performance of these signals, we define "true alarms" and "false alarms." These terms relate specifically to the "crash-alert" signals generated by our DTCAI metric (when DTCAI > 0.6). A "true alarm" is a "crash-alert" signal that occurs within a window of 1 to 20 trading days before a price peak preceding an actual historical crash (using the same 25% decline definition established in LPPL "model fit and preprocessing"). Conversely, a "false alarm" is any "crash-alert" signal outside this 1–20-day pre-crash window. These definitions assess the metric's ability to provide a timely warning within a practical window for potential action before a significant market downturn begins. Evaluating the DTCAI signals using these definitions shows that among the 199 "crash-alert" signals generated by DTCAI during the 2019–2024 period, 69 (34.7%) were classified as false alarms. This outcome contrasts sharply with the performance of the DTC metric, which generated 360 crash alerts, of which 309 (85.8%) were false alarms according to the same criteria. The substantial reduction of false alarms underscores the effectiveness of the DTC adjustment using the reliability score from the trained ANN.

Interpreting these results requires careful consideration, particularly around the COVID-19 crash in early 2020. The LPPL model and our framework are primarily designed for detecting endogenous bubble dynamics; however, the COVID-19 pandemic represents an exogenous shock. Therefore, the model may be unable to predict the precise timing of a crash triggered by such an unforeseen external event; however, Fig. 2b shows that our DTCAI metric generated "crash-alert" signals before the significant decline in February–March 2020. We posit that this reflects the framework detecting underlying market fragility or characteristics consistent with late-stage bubble dynamics that may have existed before the pandemic; the exogenous shock catalyzed the crash in an already vulnerable market. Furthermore, since the first reports of COVID-19 emerged in December 2019, the market had a period to absorb initial information before the sharp downturn began in late February 2020. This situation supports the idea that the crash was not solely a reaction to sudden news but perhaps the tipping point for pre-existing market stress. Another price drawdown after the peak in Fig. 2 occurs in 2022. The long-lasting stock price increase after the COVID-19 pandemic crash coincided with a surge in inflation, potentially attributable to the expansionary monetary policy implemented to support the economy during the pandemic (Ahn et al., 2025). However, since it peaked on January 3, the stock price experienced a regime change to a rapidly decreasing trend in early 2022. Figure 2b shows that the DTCAI metric exhibits "crash-alert" signals from late 2021, indicating that the stock price

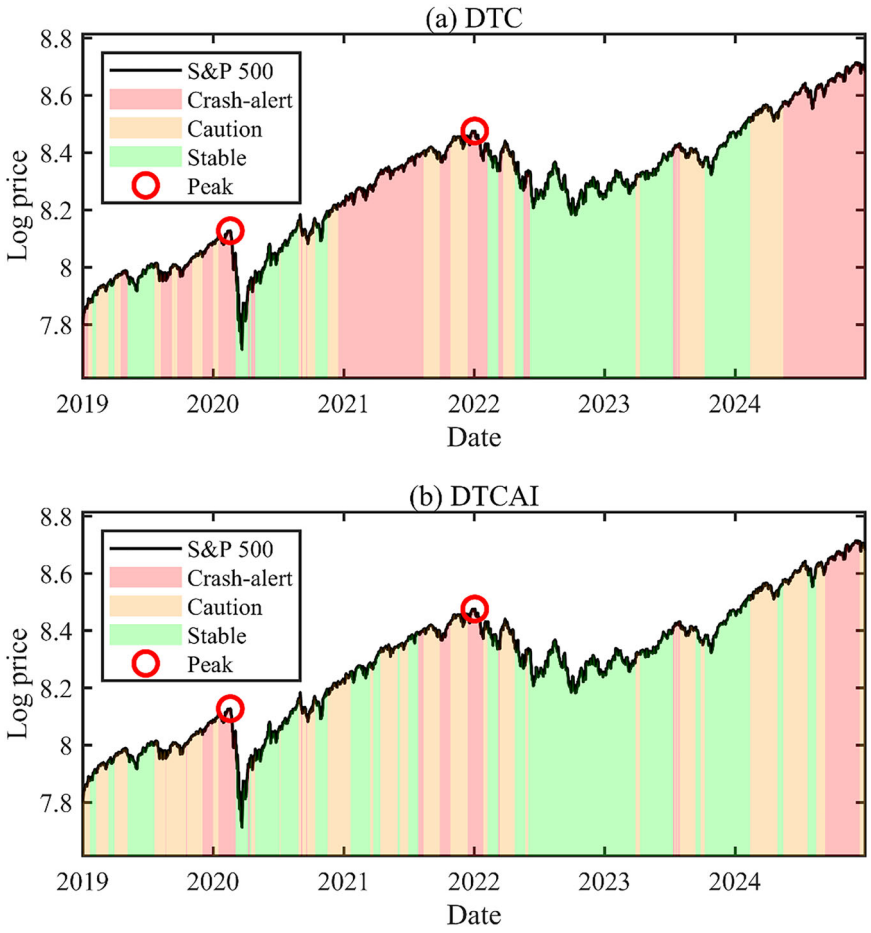

Fig. 2 Comparison of DTC and DTCAI risk metrics on the S&P 500 index. The scores of (**a**) DTC and (**b**) DTCAI. The scores of the two risk metrics (represented by the colored areas) are measured for each trading day and displayed alongside the log price of the S&P 500 index. Note the reduction in false alarms signaled by the DTCAI.

increase before the regime change was primarily due to a bubble formation. The bubble had become increasingly susceptible to a crash as time passed, which could be detected using our DTCAI metric; it burst when the government tightened the monetary policy. The COVID-19 pandemic and the policy tightening were exogenous events that triggered bubble bursts; however, the framework can provide valuable warnings of heightened risk near exogenous events if underlying vulnerabilities detectable by LPPL patterns are present. As a robustness check, Appendix B presents a further investigation into the framework's performance during the endogenous 2008 GFC crash, requiring retraining the model on pre-GFC data and Appendix C explores the framework in different markets.

Portfolio-rebalancing exercise

Experiment design. We used our proposed DTCAI risk metric to implement a dynamic portfolio-rebalancing strategy over six years (2019–2024), evaluating the metric's economic significance. We applied the risk metric as the rebalancing signal to a simple stock/bond portfolio composed of the SPDR S&P 500 ETF and the iShares Core US Aggregate Bond ETF, two widely used and liquid exchanged-traded funds (ETFs). The daily closing prices of the two ETFs from January 2019 to December 2024 are from Datastream. We accounted for transaction costs, assuming a cost of 1% applied to the value of assets traded during each rebalancing event. Dividends from both ETFs are reinvested, and the benchmark is a static buy-and-hold strategy with a 50/50 stock/bond ratio.[6]

The backtesting procedure is designed to avoid look-ahead bias. The DTCAI signal used for making rebalancing decisions on any given day t is calculated using only LPPL parameters estimated from data ending on or before day t. These are applied to the AI model, which was trained on a separate historical period preceding the 2019–2024 backtest period. The rebalancing rule is based on the pattern of signals over the preceding 21 days, ensuring that decisions rely only on past information.

The DTCAI-based portfolios adaptively rebalance asset weights based on the crash alerts ("crash-alert," "caution," and "stable") provided by our risk metric. These alerts are critical signals to adjust the portfolio's exposure to stocks and bonds as market conditions evolve and the crash risk changes. We tested three rebalancing strategies with varying risk tolerance to simulate an investor's aggressive, intermediate, and conservative risk preferences. Each strategy rebalances assets following a specific approach. The decision to rebalance is triggered after 10 "crash-alert" signals (or 10 "stable" signals) are observed within the preceding 21-day trading window. We chose 21 trading days because it closely approximates one calendar month, a conventional timeframe for portfolio review and rebalancing in both academic studies and industry practice, helping to limit portfolio turnover and associated transaction costs. Furthermore, this 21-day window is crucial to our signal confirmation process. Instead of reacting to a single daily signal, a pattern of 10 alerts within this period is required to enhance the signal's robustness and mitigate the effect of potentially noisy daily signals or isolated false alarms. The rebalancing trade is executed at the closing

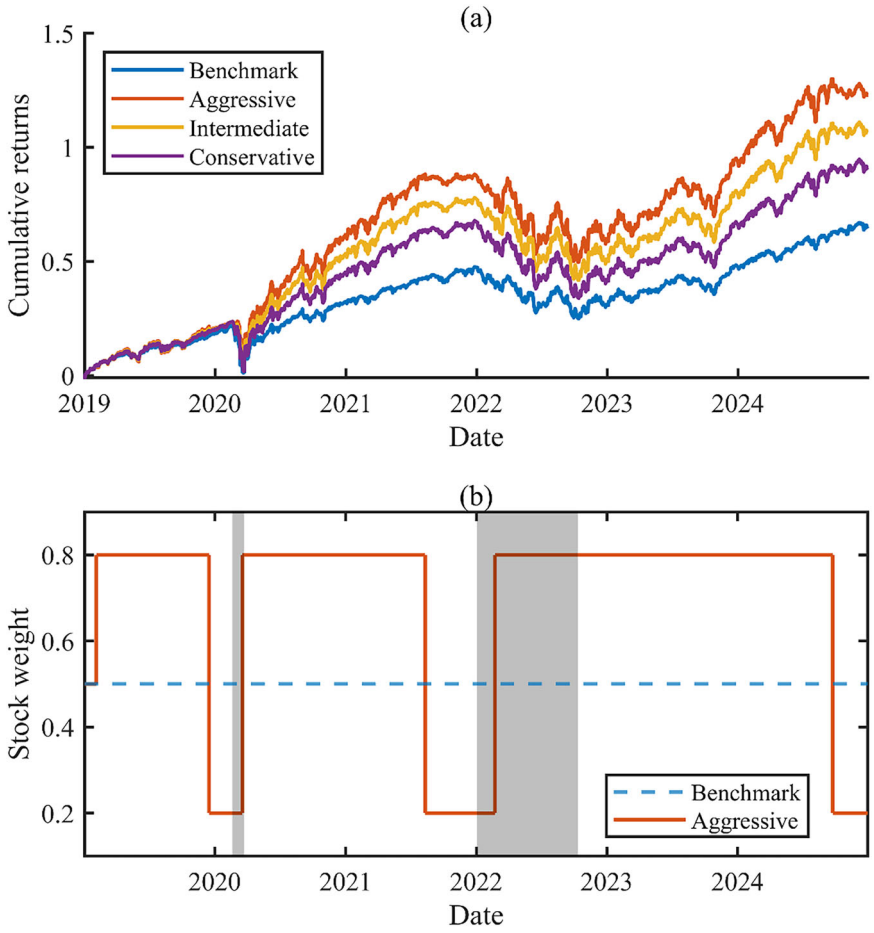

Fig. 3 Cumulative returns and portfolio allocation over time. a shows the cumulative returns, defined as the total return accrued over time. Each line is color-coded to represent a specific portfolio strategy. The aggressive strategy indicates higher risk tolerance with greater weight adjustments, the intermediate strategy reflects balanced adjustment, and the conservative strategy shows minimal adjustments to maintain lower risk. **b** shows the composition of the stock/bond of the DTCAI-based strategy (aggressive) and the benchmark by-and-hold strategy. The timings of the DTCAI signals are identical for all three strategies. The gray areas represent drawdown periods. The brief period in 2020 represents the 2020 stock market crash, where the S&P 500 index dropped by 34%. The period covering most of 2022 represents the 2022 stock market decline, where the S&P 500 index dropped by 28%.

prices of the day when this condition is met. When triggered by crash alerts, the portfolio shifts weight toward the bond for safety; when triggered by stable alerts, the weight shifts toward the stock for more market exposure.

The magnitude of the shift aligns with the investor's risk preference. Specifically, in the aggressive strategy, the asset weights fluctuate by ±0.3 from their initial allocation based on the DTCAI signal, allowing for a more significant shift in response to perceived crash risk. The intermediate strategy allows weight adjustments of ±0.2, providing a balanced response between risk management and potential returns. Finally, the conservative strategy adjusts the asset weights by ±0.1, maintaining a cautious approach to risk while allowing for dynamic rebalancing. The minimum price is monitored daily if a portfolio is rebalanced due to crash alerts. If the minimum price remains unchanged for 21 consecutive days, the asset allocation ratio is reset to its initial state.

These dynamic portfolio-rebalancing strategies aim to mitigate the risk during speculative bubbles or market downturns by shifting weights to bonds while still capturing the upside potential during periods of market growth by shifting weights to stocks. We used the 21-day observation period to mitigate the effect of false alarms, aiming to assess whether our risk metric is an effective signal for a portfolio-rebalancing strategy while considering transaction costs and maintaining reasonable limits on portfolio turnover. Finally, the three strategies allow us to reflect different risk appetites, allowing a comprehensive test of the DTCAI risk metric's practical effectiveness in real-world portfolio management.

Performance of the rebalancing strategy. The results of our portfolio-rebalancing experiments—shown in Fig. 3a and summarized in Table 5—demonstrate a clear advantage for our DTCAI-based rebalancing strategies. Across the three risk preferences (aggressive, intermediate, and conservative), our strategy consistently outperformed the benchmark regarding returns and Sharpe ratios. This outcome suggests that our risk indicator provides us with economically significant information. Figure 3b illustrates that the DTCAI ex-ante signaled the drawdowns in 2020 and 2022; therefore, investors can avoid upcoming stock price plunges and take a more aggressive position when market exposure to risks of a bubble crash is reduced through the DTCAI-based rebalancing strategies.

To further test the robustness of our results, we constructed Markowitz efficient portfolios that match the average returns of the proposed dynamic strategies for a direct comparison of the risk-adjusted performance. These portfolios were constructed using out-of-sample data spanning the entire evaluation period (2019–2024). The "target return" for each efficient portfolio presented in Table 6 was set to match the average daily return realized by the corresponding DTCAI dynamic strategy

Table 5 Descriptive statistics of daily portfolio returns (Benchmark; Stock/bond ratio of 50/50).

	Avg.	Std. dev.	Min	Max	Sharpe ratio
Benchmark	4.2711×10^{-4}	0.0076	−0.0679	0.0497	0.0440
Aggressive	6.1624×10^{-4}	0.0088	−0.0497	0.0725	0.0592***
Intermediate	5.4440×10^{-4}	0.0081	−0.0548	0.0637	0.0557***
Conservative	4.7141×10^{-4}	0.0074	−0.0600	0.0552	0.0507**

The benchmark represents a static allocation with no dynamic rebalancing, serving as a baseline for comparison. The aggressive strategy allows for the largest weight adjustments, reflecting a high-risk, high-reward approach. The intermediate strategy applies moderate adjustments to balance risk and return, while the conservative strategy employs minimal adjustments to maintain a low-risk profile. The Sharpe ratio is also included to measure risk-adjusted returns, defined as the excess return divided by the portfolio's SD. We apply the Ledoit and Wolf (2008)'s difference in the Sharpe ratio test, where the null hypothesis is that there is no difference in the Sharpe ratio of each strategy from that of the benchmark. *** and ** indicate significance at the 1% and 5% levels, respectively.

Table 6 Descriptive statistics of the Markowitz efficient portfolios.

	Target return	Stock weight	Bond weight	Sharpe ratio
Benchmark	4.2711×10^{-4}	0.6472	0.3528	0.0504
Aggressive	6.1624×10^{-4}	0.9643	0.0357	0.0508*
Intermediate	5.4440×10^{-4}	0.8440	0.1560	0.0508*
Conservative	4.7141×10^{-4}	0.7219	0.2781	0.0507

The target return for each static-efficient portfolio is set to match the average daily return of the equivalent dynamic portfolios (Table 5). The table presents each efficient portfolio's stock and bond weights and their respective Sharpe ratios. We apply the Ledoit and Wolf (2008)'s difference in the Sharpe ratio test where the null hypothesis is that there is no difference in the Sharpe ratio of each strategy from that of their respective static-efficient portfolio. * indicates significance at the 10% level.

(aggressive, intermediate, and conservative) over the same out-of-sample period. This approach allows the Sharpe ratios of static-efficient portfolios in Table 6 to reflect the differences in portfolio efficiency required to achieve realized returns equivalent to those of the dynamic portfolios. The benchmark Markowitz efficient portfolio in Table 6 has a higher Sharpe ratio (0.0504) than that of the benchmark portfolio in Table 5 (0.0440). This result is expected because, by definition, Markowitz's efficient portfolio is constructed to provide the minimum possible risk (standard deviation) for any specific level of expected return. Since the Sharpe ratio measures the portfolio's excess return per unit of risk, Markowitz efficient portfolio inherently achieves the highest Sharpe ratio when compared to other static portfolios that attain the same expected return. Conversely, a comparison of the Sharpe ratios between the Markowitz efficient portfolio and those of the dynamic DTCAI-based portfolios in Table 5 reveals that, at minimum, the aggressive and intermediate DTCAI-based portfolios exhibit higher Sharpe ratios than the static efficient portfolios. This finding suggests that aggressive and intermediate rebalancing strategies are more effective in risk reduction compared to the most efficient static portfolios.[7]

Conclusion

This study integrates AI techniques into the LPPL model, contributing to the ongoing efforts to improve the reliability of financial bubble detection and associated risk assessment. We trained three classification models (ANN, RF, and logistic regression) on a large dataset of estimated LPPL parameters derived from the largest 100 companies in the S&P 500 index over 24 years. We used these classification models to measure the reliability score of the estimated LPPL parameter of critical time and evaluated each model's ability to assign a reliability score to the critical time. Our findings show that ANN achieved the highest predictive performance among the three models. Subsequently, we introduced a novel risk indicator, DTCAI, which combines the LPPL-predicted DTC with the ANN-generated reliability score. This approach allowed us to quantify the proximity to a predicted asset price crash and its reliability. We applied DTCAI as a signal to rebalance a stock/bond portfolio to test its efficacy. Our empirical analysis shows that DTCAI-based rebalancing strategies outperform benchmark buy-and-hold strategies regardless of the investor's risk preference. In addition, our DTCAI-driven portfolios exhibit higher Sharpe ratios than efficient portfolios with matched returns, underscoring the framework's potential to enhance portfolio management.

Our AI framework extends the LPPL model's capabilities by addressing the issue related to out-of-sample forecasting. Introducing a metric that measures the reliability score for distance to endogenous crash provides an alternate view of imminent risk. This approach offers finance professionals a novel tool that can facilitate a more comprehensive risk analysis when combined with traditional risk metrics.

Despite its contributions, our framework has several limitations that warrant further research. First, the extent to which the framework can be generalized is constrained by the coverage of the training data. The AI models were trained on stocks from the S&P 500 index, which are predominantly large-cap, highly liquid US firms. Consequently, the performance of the framework may not be generalizable to less liquid or smaller-capitalization stocks. Second, the optimal classification model for the task was selected from a limited set of candidate models—namely, artificial neural networks (ANN), random forests (RF), and logistic regression—based on their prevalence in the literature and their suitability for the task. Third, the portfolio rebalancing exercise, illustrating its economic significance, employed a simplified setup. Furthermore, the proposed approach assumes that the patterns distinguishing reliable from unreliable LPPL fits remain stable over time. However, market structures may evolve due to factors such as the proliferation of algorithmic trading and regulatory changes, potentially leading to a deterioration in the model's predictive performance over extended out-of-sample periods.

Future work could mitigate these limitations by expanding the asset universe, improving the AI framework via ensemble or adaptive methods, and developing portfolio exercises that better capture real-world constraints such as rebalancing costs and complexity. Finally, this study utilizes daily closing prices, aligning with the existing LPPL literature; however, future research could investigate how this framework applies to high-frequency data. Such data might allow for detecting short-term crash signals; however, this approach would necessitate addressing the challenges related to potential changes in the LPPL

model assumptions, considering market microstructure, namely, noise. Addressing these avenues for future research could broaden the framework's scope and utility in crash prediction and risk assessment in the financial market.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 23 January 2025; Accepted: 3 September 2025;
Published online: 31 October 2025

Notes

1 The three models we selected are particularly suitable for our task because the classification task in this study is not so complicated and we do not need recent large and complex AI models such as attention-based models.

2 Once all the peaks are identified, the weight of peak i is calculated as $w_{(0,i)}$=1/(T - i), where T is the length of the time series. Then, the weights are standardized as $w_i = w_{0,i}/(\sum_{j=1}^{n} w_{0,j})$).

3 The RF model ignores temporal dependencies, leading to notorious issues in time series analyses; however, in our framework, the RF model input is a set of seven LPPL parameters, which is not a time series. LPPL parameters are estimated from the time series of stock prices and they are subsequently given as input to classification models, such as the RF model, without any time information. Therefore, our framework is free from the problem that RF models suffer from for time series data.

4 We considered alternative formulations, such as directly adjusting the predicted t_c based on the reliability score P; however, determining an appropriate functional form for such an adjustment (e.g., shifting t_c earlier or later based on P) is not straightforward and could introduce more complexity without a clear theoretical or empirical basis. The use of confidence intervals for t_c is appealing. However, generating statistically robust confidence intervals for LPPL parameters, especially t_c, is notoriously difficult and often unreliable due to the model's non-linearity and sensitivity to fitting windows.

5 The LPPL model and the proposed risk metrics can provide significant forward-looking predictions. First, we constrain parameter t_c, representing the crash time, to be larger than the end time of the subseries when we estimate the LPPL parameters, i.e., $t_c > t_2$. The subseries ends at the last observation available in practical applications of the LPPL model; thus, the LPPL model predicts the timing of an unobserved future market crash. Second, in the definition of DTC, t_c is a forward-looking parameter, and t_1 and t_2 are constant in a subseries that ends before the unobserved future market crash. Hence, DTC is always a forward-looking risk metric that measures the imminence of a future market crash; for instance, as the predicted critical time, t_c, approaches t_2 (the end of the current observation period), the DTC value approaches 1, whereas if t_c is predicted to be far in the future relative to t_2, the DTC value approaches 0. Finally, we proposed another risk metric, DTCAI, which is DTC adjusted by the reliability score of estimated LPPL model. We evaluate the reliability score of the LPPL model using AI-based classification models, which are trained using the earliest observations from our dataset. This approach ensures that the classification models do not use future information while evaluating the LPPL model's reliability for the test period.

6 We also investigated cases with other benchmarks: static portfolios with other stock/bond ratios. The results are provided in Appendix A, confirming our results' robustness.

7 The statistical significance of the static-efficient portfolio's Sharpe ratio is not central to our findings. The two key points are (1) the statistically significant outperformance of our proposed strategies (Table 5) compared to the benchmark and (2) the greater Sharpe ratios of our proposed strategies compared to those of the static-efficient portfolios.

References

Ahn K, Jang H, Kim J, Ryu I (2024a) COVID-19 and REITs crash: Predictability and market conditions. Comput Econ 63(3):1159–1172

Ahn K, Jeong M, Kim J, Tarzia D, Zhang P (2024b) Tightening policy and housing price bubbles: Examining an episode in the Chinese housing market. PLoS One 19(9):e0309483

Ahn K, Jang H, Jeong M, Sohn S (2025) The impact of futures trade on the informational efficiency of the U.S. REIT market. Financ Innov 11:45

Aldasoro I, Claudio B, Mathias D (2018) Early warning indicators of banking crises: Expanding the family. J Financ Transform 48:141–155

Andersen TG, Bollerslev T, Diebold FX, Labys P (2003) Modeling and forecasting realized volatility. Econometrica 71(2):579–625

Atiya AF (2001) Bankruptcy prediction for credit risk using neural networks: A survey and new results. IEEE Trans Neural Netw 12(4):929–935

Baker JE (1987) Reducing bias and inefficiency in the selection algorithm. Proc Second Int Conf Genet Algorithms 206:14–21

Brée DS, Joseph NL (2013) Testing for financial crashes using the log periodic power law model. Int Rev Financ Anal 30:287–297

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Bussiere M, Fratzscher M (2006) Towards a new early warning system of financial crises. J Int Money Financ 25(6):953–973

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic minority oversampling technique. J Artif Intell Res 16:321–357

Chen M (2011) Predicting corporate financial distress based on the integration of decision tree classification and logistic regression. Expert Syst Appl 38(9):11261–11272

Cont R (2001) Empirical properties of asset returns: Stylized facts and statistical issues. Quant Financ 1(2):223

Dai B, Zhang F, Tarzia D, Ahn K (2018) Forecasting financial crashes: Revisit to log-periodic power law. Complexity 2018(1):1–12

Derrida B, De Seze L, Itzykson C(1983) Itzykson C Fractal structure of zeros in hierarchical models. J Stat Phys 33:559–569

Dutta A, Bandopadhyay G, Sengupta S (2012) Prediction of stock performance in the Indian stock market using logistic regression. Int J Bus Inf 7(1):105–136

Edison HJ (2003) Do indicators of financial crises work? An evaluation of an early warning system. Int J Financ Econ 8(1):11–53

Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50(4):987–1007

Engle RF, Bollerslev T (1986) Modeling the persistence of the conditional variances. Econom Rev 5(1):1–50

Ehsani S, Linnainmaa JT (2022) Factor momentum and the momentum factor. J Financ 77(3):1877–1919

Eraker B, Johannes M, Polson N (2003) The impact of jumps in volatility and returns. J Financ 58(3):1269–1300

Fama EF, French KR (1996) Multifactor explanations of asset pricing anomalies. J Financ 51(1):55–84

Fama EF, French KR (2015) A five-factor asset pricing model. J Financ Econ 116(1):1–22

Feigenbaum JA (2001) A statistical analysis of log-periodic precursors to financial crashes. Quant Financ 1(3):346–360

Filimonov V, Sornette D (2013) A stable and robust calibration scheme for the log-periodic power law model. Phys A 392(17):3698–3707

Geraskin P, Fantazzini D (2013) Everything you always wanted to know about log-periodic power laws for bubble modeling but were afraid to ask. Eur J Financ 19(5):366–391

He H, Garcia EA (2009) Learning from imbalanced data. IEEE Trans Knowl Data Eng 21(9):1263–1284

Heston SL (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev Financ Stud 6(2):327–343

Ho TK (1995) Random decision forests. Proc 3rd Int Conf Doc Anal Recognit 1:278–282

Jang H, Ahn K, Kim D, Song Y (2018a) Detection and prediction of house price bubbles: Evidence from a new city. Lect Notes Comput Sci 10862:782–795

Jang H, Song Y, Sohn S, Ahn K (2018b) Real estate soars and financial crises: Recent stories. Sustainability 10(12):4559

Jang H, Song Y, Ahn K (2020) Can the government stabilize the housing market? The evidence from South Korea. Phys A 550:124114

Jeong M, Kim S, Yi E, Ahn K (2023) Market efficiency and information flow between the crude palm oil and crude oil futures markets. Energy Strategy Rev 45:101008

Ji G, Kong H, Kim WC, Ahn K (2020) Stochastic volatility and early warning indicator. Lect Notes Comput Sci 12137:413–421

Johansen A, Ledoit O, Sornette D (2000) Crashes as critical points. Int J Theor Appl Financ 3(2):219–255

Kurz-Kim J-R (2012) Early warning indicator for financial crashes using the log periodic power law. Appl Econ Lett 19(15):1465–1469

Laloux L, Potters M, Cont R (1999) Are financial crashes predictable? Europhys Lett 45(1):1–5

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444

Ledoit O, Wolf M (2008) Robust performance hypothesis testing with the Sharpe ratio. J Empir Financ 15(5):850–859

Lee N, Yi E, Ahn K (2020) Boost and burst: Bubbles in the bitcoin market. Lect Notes Comput Sci 12137:422–431

Lessmann S, Baesens B, Seow HV, Thomas LC (2015) Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. Eur J Oper Res 247(1):124–136

Merton RC (1976) Option pricing when underlying stock returns are discontinuous. J Financ Econ 3(1–2):125–144

Ohlson JA (1980) Financial ratios and the probabilistic prediction of bankruptcy. J Account Res 18(1):109–131

Sharpe WF (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. J Financ 19(3):425–442

Singh R, Srivastava S (2016) Stock prediction using deep learning. Multimed Tools Appl 76(18):18569–18584
Sornette D, Demos G, Zhang Q (2015) Real-time prediction and postmortem analysis of the Shanghai 2015 stock market bubble and crash. J Invest Strateg 4:77–95
Xu J, Lu Z, Xie Y (2021) Loan default prediction of the Chinese P2P market: A machine learning methodology. Sci Rep. 11(1):18759
Zhang Q, Sornette D, Balcilar M, Gupta R, Ozdemir ZA, Yetkiner H (2016) LPPLS bubble indicators over two centuries of the S&P 500 index. Phys A 458:126–139
Zhou WD, Sornette D (2006) Is there a real-estate bubble in the US? Phys A 361(1):297–308

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1A2C1A01005949, RS-2023-00217705, Taeyoung Park), and the MSIT(Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) support program (RS-2023-00259934, Taeyoung Park) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Author contributions

Ganghyeok Lee: Software, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. Minhyuk Jeong: Software, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. Taeyoung Park: Resources, Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. Kwangwon Ahn: Resources, Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing, Supervision.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-05920-7.

Correspondence and requests for materials should be addressed to Taeyoung Park or Kwangwon Ahn.