

## Research Article

# Forecasting Financial Crashes: Revisit to Log-Periodic Power Law

Bingcun Dai,<sup>1</sup> Fan Zhang,<sup>2</sup> Domenico Tarzia ,<sup>3</sup> and Kwangwon Ahn <sup>4</sup>

<sup>1</sup>Saïd Business School and Green Templeton College, University of Oxford, Park End Street, Oxford OX1 1HP, UK

<sup>2</sup>Shenzhen Goofar Sanxin Fund Management, Qiushi Building, Shenzhen 518000, China

<sup>3</sup>HSBC Business School, Peking University, University Town, Shenzhen 518055, China

<sup>4</sup>Graduate School of Future Strategy, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

Correspondence should be addressed to Kwangwon Ahn; [k.ahn@kaist.ac.kr](mailto:k.ahn@kaist.ac.kr)

Received 19 February 2018; Revised 19 April 2018; Accepted 5 June 2018; Published 1 August 2018

Academic Editor: Gonzalo Ruz

Copyright © 2018 Bingcun Dai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We aim to provide an algorithm to predict the distribution of the critical times of financial bubbles employing a log-periodic power law. Our approach consists of a constrained genetic algorithm and an improved price gyration method, which generates an initial population of parameters using historical data for the genetic algorithm. The key enhancements of price gyration algorithm are (i) different window sizes for peak detection and (ii) a distance-based weighting approach for peak selection. Our results show a significant improvement in the prediction of financial crashes. The diagnostic analysis further demonstrates the accuracy, efficiency, and stability of our predictions.

## 1. Introduction

Financial crises that follow asset price bubbles have been observed in various markets throughout history. Bubbles refer to asset prices that exceed the fundamental values based on supply and demand. Skewed asset prices fail to reflect the fundamentals well; thus, in turn, they may have an important effect on resource allocations [1]. Moreover, the bursting of a bubble, for example, a dramatic collapse of the stock market, may bring the economy into an even worse situation, such as the great recession, and dysfunction in the financial system, which proves the importance of understanding asset price bubbles.

Recently, a lot of attempts have been made to introduce bubbles into asset pricing models. Predominantly, two streams of theoretical frameworks shed light on this issue, that is, rational models and behavioural models. The rational models optimise the behaviour of a representative agent with complete processing of information; however, bubbles may still exist due to market imperfection like information asymmetry and short sale constraints [2]. The behavioural models, a different view, are characterised by a framework of heterogeneous agents in which at least some agents in the economy are bounded as rational and others as irrational. In behavioural

models, bubbles arise due to either limited arbitrage or heterogeneous beliefs [3]. Although empirical evidence supports the validation of these models, none of them provides the space for predicting the critical time of financial bubbles with a substantial significance.

As an alternative to explaining bubbles, a framework called the log-periodic power law (LPPL) model has gained a lot of attention with the many successful predictions it made [4–6]. Johansen et al. [4] proposed the LPPL model, which assumes that there exist two types of agents in the market: a group of traders with rational expectations and a group of noise traders with herding behaviour. The noise traders are organised into networks, and they tend to imitate others. At the macrolevel, all the agents will continue investing where arbitrage is limited because rational traders lack the knowledge about the time of crash and are assumed to be risk-neutral. It is still rational for them to invest on speculative assets because the risk of a crash is compensated for by the profits. As a consequence, rational traders will self-limit their arbitrage behaviour. The herding behaviour of noise traders is the origin of the positive feedback process; that is, given a high price, the imitation among noise traders leads to increased demand, which pushes the price further up.

The empirical literature has employed a variety of approaches to estimate the LPPL model. Johansen et al. [4] first used the so-called *Tabu search* and the Levenberg–Marquardt algorithm (LMA) to deduce an evolution law for stock prices before the crashes in the United States and Hong Kong. Johansen and Sornette [7] identified and analysed 21 significant bubbles followed by large crashes or severe collapses and found that the LPPL adequately describes speculative bubbles in emerging markets. Liberatore [8] introduced a price gyration method combined with the LMA to predict the critical time of financial bubbles for the Dow Jones Industrial Average (DJIA) and the Standard & Poor’s 500 Index (S&P 500). Pele [9] proposed an extension of the approach of Liberatore [8] using time series peak detection and predicted the Bucharest Exchange Trading-Investment Fund (BET-FI) crash in 2007. Kurz-Kim [10] applied the LPPL to detect the stock market crash in Germany, which demonstrated that the LPPL could be used for constructing an early warning indicator for financial crashes. Geraskin and Fantazzini [11] presented alternative methodologies for diagnostic tests and graphical tools to investigate the gold bubble in 2009. Korzeniowski and Kuroпка [12] used a genetic algorithm (GA) to fit the LPPL based on the DJIA and WIG20 (the WIG20 is a stock market index of the 20 largest joint-stock companies listed on the Warsaw Stock Exchange) time series and found it to be useful as a forecasting tool for financial crashes.

However, the research still has encountered problems when forecasting the critical time of financial bubbles with LPPL. First, prediction results are sensitive to the initial values because the optimisation algorithms based on derivatives, for example, gradient and curvature, can easily be trapped in local minima. Second, although the problem of local minima can be overcome by using a nonlinear optimisation algorithm, for example, GA, which requires no information about solution surface, we still do not have a proper way of providing a reasonable initial population. Finally, there has been neither sufficient diagnostic analysis for the LPPL model nor a thorough assessment of its goodness-of-fit.

Our study improves the estimation method for the LPPL model. To the best of our knowledge, we are the first to generate an initial population for GA using information from historical data. The method to create an initial population is called improved price gyration, which consists of three steps. First, detect peaks which are local maxima within a window (a periodic cycle). In this paper, the window size is not fixed, which allows for the possibilities of different lengths of the cycle. Second, select three consecutive peaks that were detected in the previous step. The more recent the peak, the greater the probability it will be selected, which we call a distance-based weighting approach. Third, given the fact that consecutive peaks are also consecutive in angle with an interval of  $2\pi$ , we can derive three parameters for the LPPL model (see Section 3 for details).

Our extensive approach avoids being trapped in local minima and provides a good and robust forecast, with the imposition of constraints on LPPL parameters. The results show that our algorithm outperforms with regard to capturing financial crashes. Our predictions of critical times are

highly concentrated around the actual times when crashes took place. Using diagnostic analysis, we also show relatively small and stationary residuals.

The remainder of this paper is organised as follows: Section 2 summarises the underlying mechanism of the LPPL model. Section 3 explains the methodology and data that we use to fit the LPPL parameters. Section 4 presents the prediction results, diagnostic tests, and model comparisons. In the final section, we provide a conclusion and discuss future research.

## 2. Theoretical Framework

There are two groups of traders, a group of rational traders who are identical in their preferences and characteristics and a group of noise traders whose herding behaviour leads to bubbles. The rational traders, however, do not likely eliminate mispricing through arbitrage but continue investing as noise traders do because the time of a crash is unknown and the risk of a crash is compensated for by a higher return generated by bubbles. Therefore, the no-arbitrage condition resulting from the rational expectation theory is more than a useful idealisation, as it describes a self-adaptive dynamic state of the market [4].

The dynamics of the asset price  $p_t$  before a crash is given by the following stochastic differential equation

$$dp_t = \mu_t p_t dt - \kappa p_t dj, \quad (1)$$

where  $\mu_t$  is the drift,  $j$  is the jump process whose value is zero before the crash and one after the crash, and  $\kappa$  is the jump size.

We assume no arbitrage in the market so that the price process satisfies the martingale condition  $E_t[dp_t] = \mu_t p_t dt - \kappa p_t h_t dt = 0$ , where  $h_t$  denotes the hazard rate (hazard rate is the probability that if the bubble survives to a certain point, it will crash during the next unit of time). Then, we have

$$\mu_t = \kappa h_t. \quad (2)$$

This indicates that rational traders are willing to accept the crash risk only if they are rendered by high profits, which is a risk-return trade-off. Substituting (2) into (1), we obtain the differential equation before the crash given by  $d(\ln p_t) = \kappa h_t dt$ , whose solution is

$$\ln \left( \frac{p_t}{p_{t_0}} \right) = \kappa \int_{t_0}^t h_s ds. \quad (3)$$

The macrolevel hazard rate can be explained in terms of the microlevel behaviour of noise traders. A large amount of simultaneous sell-off, which triggers a crash, is attributed to the noise traders’ tendency to imitate their nearest neighbours. Besides the tendency to herd as one force that tends to create order, an idiosyncratic signal is received as the other force to influence noise traders’ decisions, which causes disorder to fight with imitation. A crash happens when order wins, while disorder dominates before the critical time.

Johansen et al. [4] introduced a dynamic stochastic model for microlevel behaviour in which each trader  $i$  (for  $i = 1, \dots, n$ ) can either buy ( $s_i = +1$ ) or sell ( $s_i = -1$ ). The current state of trader  $i$  is determined by

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

where  $K$  is the tendency toward imitation,  $N(i)$  is the set of traders who influence trader  $i$ ,  $\sigma$  is the tendency toward idiosyncratic behaviour, and  $\varepsilon_i$  is a random variable.

Note that (4) only describes the state of an individual agent. To explain the macrolevel hazard rate, we should be concerned with the average state of the whole system. In particular, the hazard rate can be represented by the sensitivity of the average state to a global influence, which is called the susceptibility of the system. Assume that a global influence term  $G$  is added to (4)

$$s_i = \text{sign} \left( K \sum_{j \in N(i)} s_j + \sigma \varepsilon_i + G \right), \quad (5)$$

and the average state of the market is given by  $M = (1/I) \sum_{i=1}^I s_i$ . Then, the susceptibility of the system is defined as  $X = dE[M]/dG|_{G=0}$ , which measures the sensitivity of  $M$  to a small change in the global influence. The form of  $X$  depends on the structure of the network that links individual agents.

Johansen et al. [4] proposed a hierarchical diamond lattice structure to model the network. The lattice structure has a general solution of susceptibility given by a first-order expansion as follows

$$X \approx B_0 (K_c - K)^{-\gamma} \{1 + C_0 \cos [\omega \ln (K_c - K) + \phi]\}, \quad (6)$$

where  $K_c$  is the tendency toward imitation at the critical time  $t_c$ . If  $K$  evolves smoothly, we can apply a first-order Taylor expansion around the critical point  $t_c$ . Then, prior to  $t_c$ , we get the following approximation

$$K_c - K \propto (t_c - t). \quad (7)$$

Given that the hazard rate of the crash behaves in the same way as the susceptibility in the neighbourhood of the critical point, we get

$$h_t \approx B_1 (t_c - t)^{-(1-\beta)} \{1 + C_1 \cos [\omega \ln (t_c - t) + \phi]\}, \quad (8)$$

where  $1 - \beta > 0$  is the critical exponent of hazard rate like that of susceptibility.

Substituting (8) into (3) and integrating provide (9), which is known as the LPPL model

$$Y_t = A + B(t_c - t)^\beta \{1 + C \cos [\omega \ln (t_c - t) + \phi]\}, \quad (9)$$

where  $Y_t > 0$  is the log of price at time  $t$ ,  $A > 0$  is the log-price at the critical time  $t_c$ ,  $B < 0$  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 oscillations around the exponential trend,  $t_c > 0$  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 bubble, and  $\phi \in [0, 2\pi]$  is a phase parameter.

### 3. Methods and Data

**3.1. Fitting the LPPL Parameters.** The basic form of the LPPL given by (9) requires the estimation of seven parameters. The parameter set must be such that the root mean square error (RMSE) between the observation and the predicted value of the LPPL model is minimised as follows

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - Y_t)^2}, \quad (10)$$

where  $y_t$  and  $T$  denote the log of observation at time  $t$  and the number of trading days in the dataset.

Let  $f_t = (t_c - t)^\beta$  and  $g_t = (t_c - t)^\beta \cos [\omega \ln (t_c - t) + \phi]$ ; then, the LPPL model, (9), can be rewritten as

$$Y_t = A + Bf_t + BCg_t. \quad (11)$$

It is straightforward to estimate the parameters  $A$ ,  $B$ , and  $C$  using ordinary least squares (OLS) given the four parameters  $\beta$ ,  $\omega$ ,  $t_c$ , and  $\phi$ . Thus, estimating these four parameters is the main task for applying the LPPL model.

We fit the LPPL parameters with two steps. In the first step, we produce the initial values for the parameters with a price gyration method. In the second step, we optimise these parameters using a nonlinear optimisation algorithm, GA. Such an indirect method is acceptable for nonlinear models, because the estimation results are sensitive to initial values and stochastic environment (the advantage of a two-step approach is, namely, from price gyration, that we can obtain an initial population for the parameters, purely based on data. However, it takes a long computing time, which is a clear disadvantage. It will not be a big issue due to the recent developments in parallel computing and multiprocessing (see Appendix A)).

Liberatore [8] defined a price gyration method to produce the initial values of the LPPL parameters by visually inspecting stock prices as follows:

- (1) Identify three consecutive stock price peaks, that is,  $i$ ,  $j$ , and  $k$ ;
- (2) Estimate the initial values of  $t_c$ ,  $\omega$ , and  $\phi$  from the price gyration as  $t_c = (\rho k - j)/(\rho - 1)$ ,  $\omega = 2\pi/\ln \rho$ , and  $\phi = \pi - \omega \ln (t_c - k)$  with  $\rho = (j - i)/(k - j)$ ;
- (3) Set the initial values of other two parameters, that is,  $\beta = 1$  and  $C = 0$ ; and
- (4) Estimate the initial values of  $A$  and  $B$  using an OLS fit

$$y_t = A + B(t_c - t) + \varepsilon_t. \quad (12)$$

Pele [9] extended Liberatore's [8] approach using an automatic peak detection algorithm [13] described as follows:

- (1) Define a peak function  $S_i$  which associates a score with element  $y_i$  and distance  $\kappa$

$$S_i[\kappa, y_i] = \frac{1}{2} \left\{ \max [y_i - y_{i-1}, \dots, y_i - y_{i-\kappa}] + \max [y_i - y_{i+1}, \dots, y_i - y_{i+\kappa}] \right\}; \quad (13)$$

- (2) Screen the series of  $S_i$  using  $S_i > 0$  and  $S_i - m > hs$ , where  $m$  and  $s$  are the mean and standard deviation of  $S_i$  and  $h$  is a positive coefficient (the value of  $h$  indicates how many standard deviations the selected peaks should be away from the mean. A higher  $h$  value means a stricter rule on detecting peaks.  $h$  is typically set within  $1 \leq h \leq 3$ . In our paper, we follow a moderate rule on peak detection and choose  $h = 1.5$  following Palshikar [13]); and
- (3) Then, retain only one peak with the largest value from any set of peaks within a fixed distance  $\kappa$  which rolls across the whole sample, and finally obtain the peak series.

Once the peaks are detected, price gyration might encounter following problems. The prediction results are not stable for different window sizes, and the estimation of critical times is not sufficiently accurate if peaks are too far from the last day of observation. To eliminate these issues, we relax and improve the idea of a fixed window size and equally weighted peaks. Our window size  $\kappa$  for peak detection is no longer fixed, which allows us to test for different possibilities of a fluctuating cycle of LPPL growth. Because more recent data have included more information on forecasting, we further implement a distance-based weighting approach for peak selection. After obtaining a series of peaks, we calculate the weight of each peak  $i$

$$w_{0,i} = \frac{1}{T-i}, \quad (14)$$

where  $T$  is the sample size of the time series. We standardise the value so that the sum of the weights of all the peaks is equal to 1. Then, the weight of each peak  $i$  is

$$w_i = \frac{w_{0,i}}{\sum_{j=1}^n w_{0,j}}, \quad (15)$$

where  $n$  is the total number of peaks.

The second cornerstone of our algorithm is the use of a GA to fit the LPPL model. Compared with other nonlinear optimisation algorithms, such as the quasi-Newton and the LMA, the GA has many advantages. It avoids potential local minima because the search for solution runs in parallel and does not require additional information about the shape of the calculated plane. Moreover, the objective function does not need to be continuous or smooth. The GA is implemented using the following steps:

- (1) Each member of the initial population is a vector of the seven LPPL parameters ( $A$ ,  $B$ ,  $C$ ,  $t_c$ ,  $\beta$ ,  $\omega$ , and  $\phi$ )

TABLE 1: Restriction on LPPL parameters.

Parameter	Constraint	Literature
$A$	$(\max(P), +\infty)$	Korzeniowski and Kuroпка [12]
$B$	$(-\infty, 0)$	Lin et al. [15]
$C$	$(-1, 1)$	Lin et al. [15]
$t_c$	$(t, \infty)$	Korzeniowski and Kuroпка [12]
$\beta$	$[0.1, 0.9]$	Lin et al. [15]
$\omega$	$[4.8, 13]$	Johansen [16]; Lin et al. [15]
$\phi$	$[0, 2\pi]$	Lin et al. [15]

generated by our improved price gyration algorithm. The RMSE is calculated for each member;

- (2) An offspring is produced by randomly drawing two parents, without replacement, and calculating their arithmetic mean. If any parameter value is outside the constraint, it is set as the closest boundary value;
- (3) A mutation perturbs the solution so that new region of the search space can be explored. The mutation process is performed by adding a perturbation variable to each coefficient in the current population. If the perturbation drives the parameters out of the constraints, the closest boundary value will be given to these parameters as in step 2;
- (4) After breeding and mutation, we merge the newly generated individuals into the population. All the solutions are ranked according to their RMSEs in an ascending order, and only half of the best solutions can survive to the next generation; and
- (5) We iterate this procedure and choose the best fit as the final solution.

Johansen and Sornette [7] and Jacobsson [14] found that whether an LPPL model can capture crashes well depends, to some extent, on the specific bounds of the critical parameters  $\beta$  and  $\omega$ . Based on their finding, we impose constraints on the LPPL parameters that are consistent with the previous literature. Table 1 defines the constraints on the LPPL parameters.

Our algorithm combines the price gyration method with the constrained GA, which extends existing research using a floating window size for peak detection and the distance-based weighting approach for peak selection. To summarise, the algorithm which we propose is implemented as the following steps:

- (1) Detect the peaks of the sample with window size  $\kappa$  following Palshikar [13] and Pele [9];
- (2) Assign the distance-based weight to each peak;
- (3) Randomly select three consecutive peaks based on the weights;
- (4) Use these three consecutive peaks for price gyration and obtain the initial values for  $t_c$ ,  $\omega$ , and  $\phi$ ;

- (5) Set the initial values  $\beta = 1$  and  $C = 0$ , and estimate the initial values of  $A$  and  $B$  using OLS;
- (6) Repeat steps 3 to 5, and obtain a series of initial values for the seven LPPL parameters;
- (7) Find the LPPL parameters using the constrained optimisation, GA, with the initial population of the parameters from step 6; and
- (8) Repeat steps 1 to 7 with changing the window size  $\kappa$ , and obtain the prediction interval for the critical time  $t_c$ .

3.2. *Data.* First, we need to identify financial bubbles and crashes. A financial bubble occurs when the asset price continues to increase for a long period of time beyond its fundamental value, whereas a financial crash is defined as a substantial decrease of the asset price when the bubble bursts. A critical time is the peak of a bubble that initiates a crash. Following Brée and Joseph [17], we identify a peak based on two conditions: (i) 262 weekdays *prior* to the peak, there is no value higher than the peak; and (ii) *after* the peak, stock prices decrease by more than 25% within a period of 60 weekdays.

Based on these criteria, we choose four stock market bubbles and market crashes that occurred during different time periods and in different financial markets: (i) the dot-com bubble in the late 1990s (stock markets saw their equity value increased rapidly from growth in the Internet sector and related fields in the late 1990s. The collapse of the bubble occurred during the period of 1999–2001); (ii) the Japanese asset price bubble in the late 1980s (real estate and stock prices were greatly inflated. The bubble was characterised by a rapid acceleration of asset prices and overheated economic activity, in addition to an uncontrolled money supply and credit expansion. By August 1990 (the fifth monetary tightening by the Bank of Japan), the NIKKEI stock index had plummeted to half of its peak price); (iii) the Hong Kong stock market bubble of 2007 (the announcement of the “through train” scheme by the State Administration of Foreign Exchange (SAFE) caused a frenzied boom in the Hong Kong stock market in 2007. In less than two months, the Hang Seng Index (HSI) increased from 20,000 to a peak of 31,958 at the end of October. As the central government of China postponed the “through train” scheme indefinitely, the Hong Kong stock market saw a substantial daily decrease of 1,526 points on December 5, 2007); and (iv) the Chinese stock market bubble of 2015 (the Shanghai Composite Index (SSEC) increased by approximately 150% within a year. Because of a series of government policies to deleverage the stock market, a third of the value of A-shares on the Shanghai Stock Exchange was lost within a month after June 2015. Subsequently, the market underwent two further crashes, in August 2015 and January 2016). The percentage decreases of four indices within 60 weekdays are 37% (NASDAQ), 28% (NIKKEI), 31% (HSI), and 43% (SSEC), respectively. Table 2 summarises information about these

TABLE 2: Data description. “Time period” is the maximum time span used in our estimation, and “Observation” is the total number of data points in the sample. “Critical time” is the date which corresponds to the peak of bubbles.

Index	Time period	Observation	Critical time
NASDAQ	08/10/1998–10/02/2000	339	10/03/2000
NIKKEI	11/11/1987–01/12/1989	507	29/12/1989
HSI	17/05/2004–02/10/2007	846	30/10/2007
SSEC	27/06/2013–15/05/2015	460	12/06/2015

TABLE 3: Summary of predicted critical times. The last date is divided into three different groups, which are from one to three months before the actual critical point.

Index	Last date	95% prediction interval	IQR	$P_{60}$
NASDAQ	10/02/2000	[13/01/2000–17/05/2000]	25	99%
	12/01/2000	[23/12/1999–01/05/2000]	43	99%
	14/12/1999	[14/12/1999–12/04/2000]	13	98%
NIKKEI	01/12/1989	[18/12/1989–07/05/1990]	19	86%
	01/11/1989	[27/11/1989–22/03/1990]	10	99%
	03/10/1989	[23/10/1989–18/01/1990]	11	100%
HSI	02/10/2007	[09/10/2007–12/02/2008]	19	97%
	04/09/2007	[02/10/2007–02/01/2008]	18	99%
	07/08/2007	[03/09/2007–10/12/2007]	17	99%
SSEC	15/05/2015	[17/04/2015–29/09/2015]	17	97%
	16/04/2015	[18/03/2015–14/09/2015]	34	93%
	18/03/2015	[04/02/2015–14/08/2015]	10	77%

four episodes. The daily closing prices of the four indices are from the WIND database.

After identifying the financial bubbles and crashes, we need to select carefully the time window to estimate the LPPL model. Following Johansen and Sornette [7] and Brée and Joseph [17], we select the forecasting window as follows: (i) the time window starts at the end of the previous crash, that is, the lowest point since the last crash (as we select the most significant crash for each index, the amplitude of the previous crash cannot be as large as the major crash that we are investigating (i.e., more than a 25% decrease within a period of 60 weekdays), and thus the choice of the starting point is disputable. In this paper, we consistently choose the previous crash according to the following criteria: (i-1) when there is no higher value than the peak during a period of 262 weekdays prior to the peak and (i-2) when stock prices decrease by more than 20% within a period of 60 weekdays after the peak. As a result, prior to the major crash, there are less severe collapses in the selected four markets, which declined by 29% (NASDAQ), 21% (NIKKEI), 21% (HSI) and 23% (SSEC), respectively, within 60 weekdays); (ii) the day with the peak value of the index is the actual critical point of the bubble; and (iii) the last day of observation is divided into three different groups, which are ranged from one to three months (one quarter) before the actual critical time.

TABLE 4: LPPL parameters of the best fit.

Index	$A$	$B$	$C$	$\beta$	$\omega$	$t_c$	$\phi$	RMSE
NASDAQ	8.5382	-0.0059	-0.1787	0.8771	5.0295	406.3596	2.7165	0.0415
NIKKEI	10.6084	-0.0026	-0.1057	0.8430	5.8084	559.0452	5.2227	0.0239
HSI	10.1877	-0.0175	-0.0527	0.5663	5.1246	846.0001	4.6871	0.0290
SSEC	8.7598	-0.0083	0.1961	0.7985	5.1113	638.0736	1.5897	0.0379

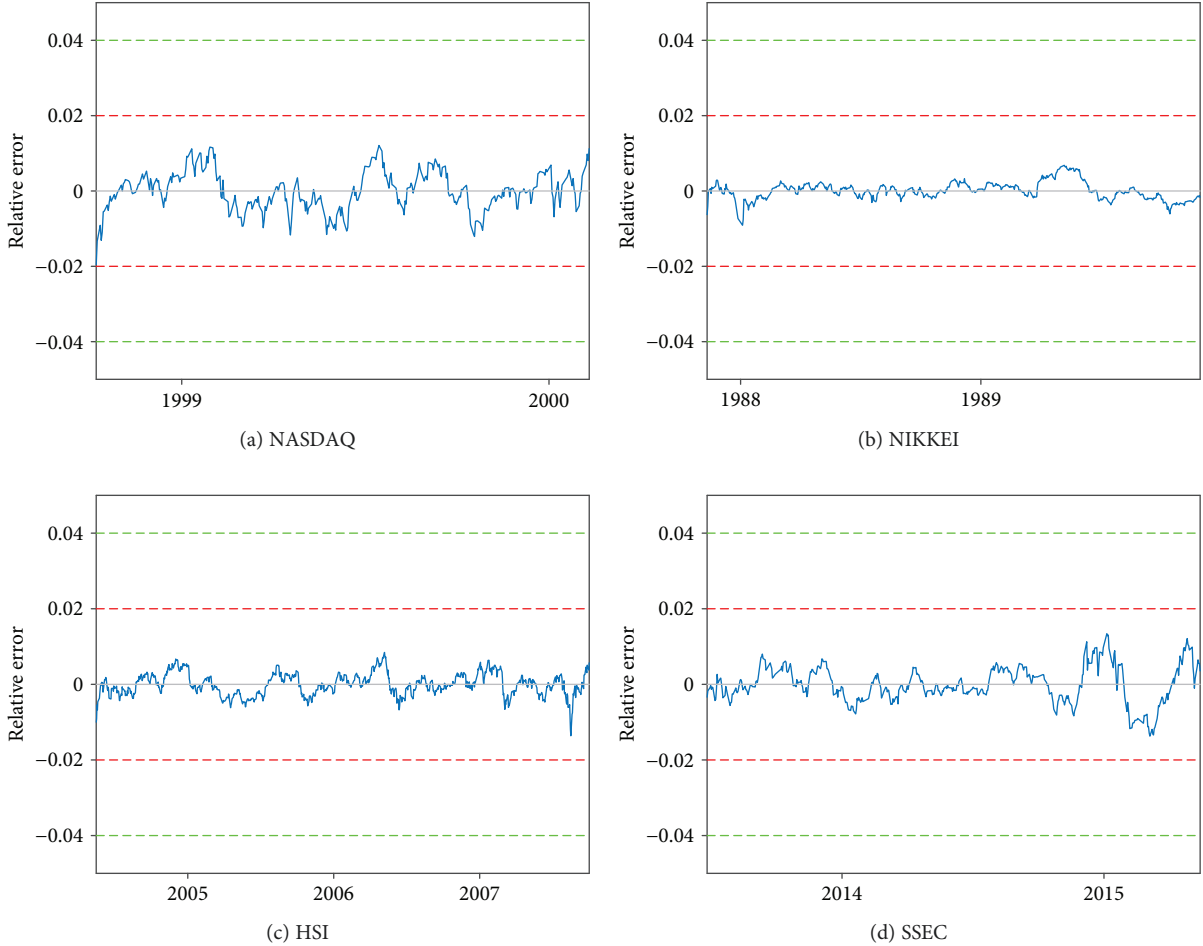


FIGURE 1: Relative error analysis of the best LPPL fits.

## 4. Results and Discussion

**4.1. Empirical Results.** We apply our algorithm to predicting the critical times of financial bubbles in the aforementioned four episodes. Table 3 summarises the forecasting results of our algorithm. The last column  $P_{60}$  denotes the percentage of predicted critical times within 60 weekdays around the actual critical time. IQR is the interquartile range of the predicted critical times, which is computed by subtracting the first quartile from the third quartile. Table 3 shows that our algorithm provides a good forecast of the four stock market crashes.

We forecast a 95% prediction interval of critical times with a period of no more than 100 weekdays from the actual critical time for all the stock market crashes. The interquartile

TABLE 5: Unit root test for LPPL residuals. We reject the null hypothesis; that is, the residuals have a unit root, at 1% significance level (\*\*\*)

Index	ADF	PP
NASDAQ	-3.88***	-4.60***
NIKKEI	-3.43***	-4.22***
HSI	-4.69***	-5.31***
SSEC	-3.54***	-3.66***

ranges of the predicted critical times are all within 50 weekdays, which means that our predictions are highly concentrated. Even though the last observation of the sample

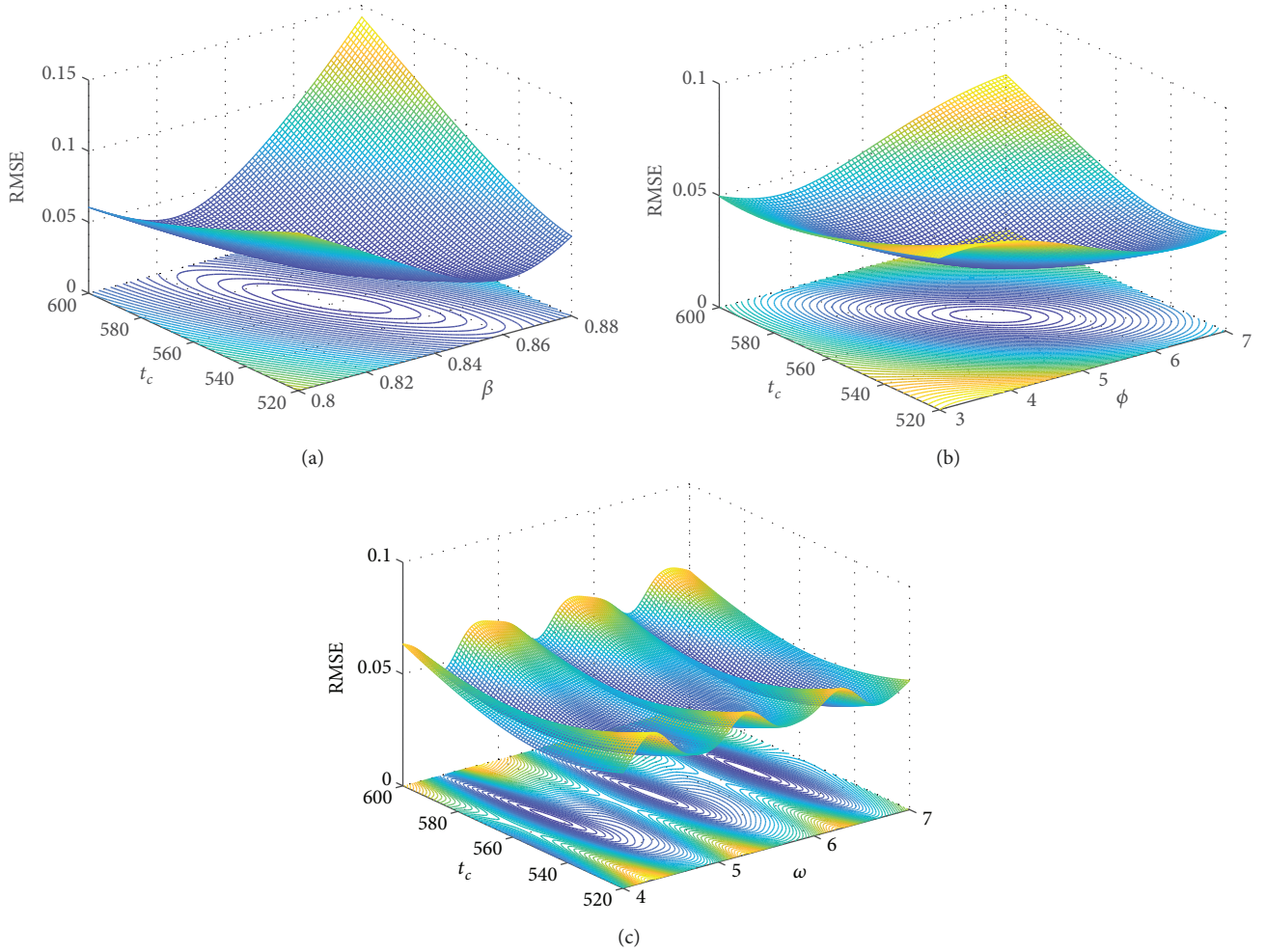


FIGURE 2: Sensitivity analysis of the RMSE to LPPL parameters (NIKKEI).

changes from one month to three months before the actual critical time, our prediction intervals demonstrate a stable behaviour. Most of the predicted critical times are well within 60-weekday horizon of the actual critical time for all four episodes ( $P_{60}$ ).

In addition to forecasting the critical time, it is important to justify the LPPL calibration, that is, to check the stylised features of LPPL, mentioned in Table 1. Table 4 shows the best LPPL fits for the four financial crashes estimated by our algorithm. Since we use the constrained GA, all the estimated parameters are well within the boundaries. The two conditions  $B < 0$  and  $0.1 \leq \beta \leq 0.9$  ensure a faster-than-exponential acceleration of the log price, with a vertical slope at the critical time  $t_c$  [15]. The positive hazard rate always holds because the absolute value of  $C$  is restricted in one unit. The value of  $\omega$  corroborates existing studies such as Johansen [16], who found that  $\omega \approx 6.36 \pm 1.56$  for 30 crashes in major financial markets.

**4.2. Diagnostic Tests.** To reduce the possibility of false alarms, it is necessary to conduct a diagnostic analysis to demonstrate our predictions. We conduct the diagnostic analysis by considering the relative errors, the unit root test of the

LPPL residuals, the sensitivity analysis of the LPPL parameters, and a crash lock-in plot (CLIP) analysis.

The relative error analysis of the best fits shows that the bubbles are well captured by our algorithm. In the analysis by Johansen et al. [4], most of the relative errors of the best fits were below 5%. Our algorithm significantly improves their results, demonstrating more accurate performance in capturing the four financial bubbles. The significance of our findings is quite evident and immediate in Figure 1, where the relative errors of all the fitting points of these four indices are well below 2%.

One key property of the LPPL model is that the residuals follow a mean-reverting Ornstein-Uhlenbeck (OU) process. Table 5 shows the results of the unit root tests. Both the augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests with two lags for the best fit of the four indices reject the null hypothesis at a 1% significance level, which means that the residuals do not have a unit root but are stationary and thus compatible with a mean-reverting OU process.

We also investigate how sensitive the RMSE is to the variations of the LPPL parameters. Because  $A$ ,  $B$ , and  $C$  are always estimated given the four parameters  $\beta$ ,  $\omega$ ,  $t_c$ , and  $\phi$ , we examine the sensitivity of the LPPL fit to the variations

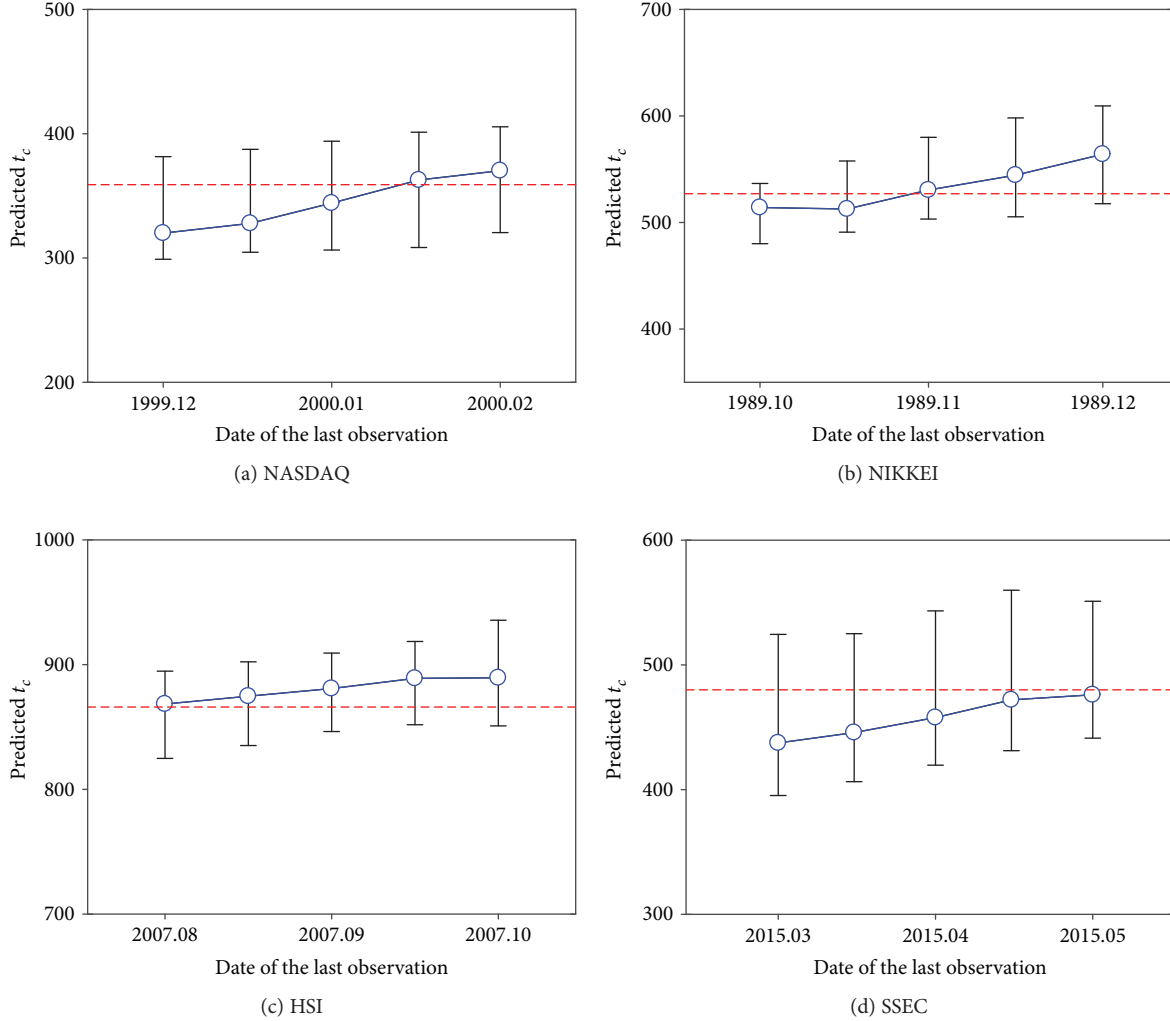


FIGURE 3: Crash lock-in plots with rolling estimation windows.

in these four parameters. We let  $t_c$  and one parameter among  $\beta$ ,  $\omega$ , and  $\phi$  vary for the NIKKEI, while the remaining parameters are fixed. The results in Figure 2 show that the variation of the RMSE is relatively smooth with respect to  $\beta$ ,  $t_c$ , and  $\phi$  and that of the RMSE is highly sensitive to small fluctuations in  $\omega$ . Figure 2 provides significant evidence supporting the choice of a GA instead of the LMA to fit the LPPL model. If the initial value of  $\omega$  is close to a local minimum, the searching algorithm of the LMA can be easily trapped because it achieves a local optimum. Compared with the LMA, the search of a GA runs in parallel, which means that even when a local minimum has been found, small variations in the parameters might avoid the search procedure becoming trapped. In addition, it is natural to obtain multiple values of  $\omega$  leading to a local minimum of RMSE.  $\omega$  is the angular frequency for a periodic function. A periodic function may have multiple periods, as  $f(t) = f(t + nT)$ , where  $T = 2\pi/\omega$  is the shortest period for  $n = 1, 2, \dots$ , and so on. The estimation result of  $\omega$  in Table 4 is in line with Johansen [16]. The log-periodic oscillations

can be neither too slow (otherwise they would contribute to the trend) nor too fast (otherwise they would fit the random component of the data) [15].

Finally, we analyse CLIP following Fantazzini [18]. A CLIP is a useful tool for tracking the development of a bubble and understanding whether a possible crash is imminent. The main idea of a CLIP is to plot the date of the last observation in the estimation sample on the horizontal axis and the estimated crash date  $t_c$  on the vertical axis. If a regime change in the stock market is approaching, then, the estimated  $t_c$  should be stabilised around a constant value close to the critical time. To implement a CLIP, we continue changing the last observation of our estimation sample consecutively from one month to three months before the actual critical time and make predictions for the four financial bubbles. From Figure 3, we can see that our results for  $t_c$  are stable, especially from the last one quarter before the crash. Moreover, the predicted critical times  $t_c$  are very close to the actual critical time. These results indicate that when a crash is imminent, our algorithm provides a robust and precise forecast of the critical time.



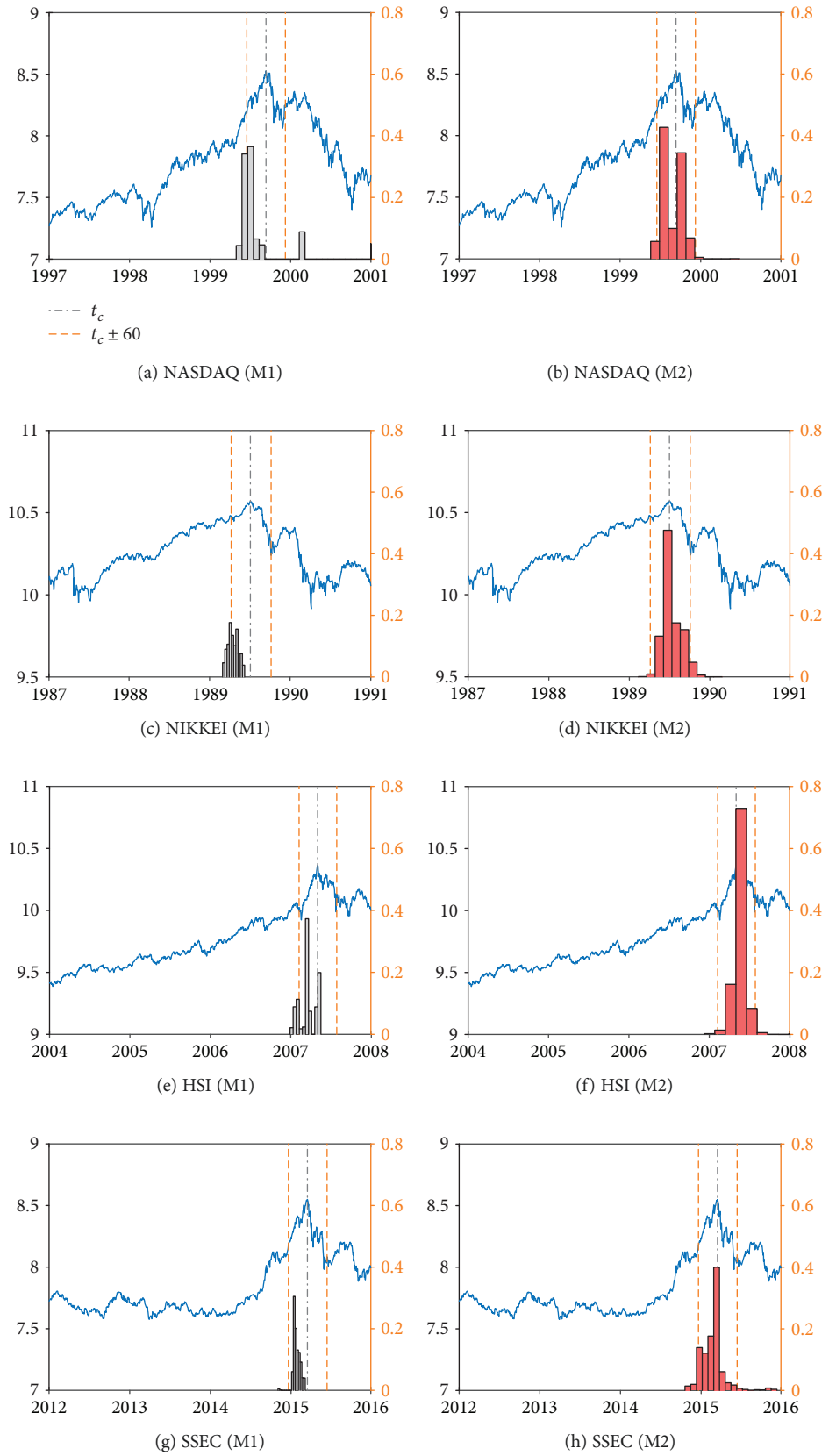


FIGURE 4: Logarithm of the stock prices and corresponding alarms.

**4.3. Model Comparison.** As a final step, we compare the prediction accuracy of critical time  $t_c$  for the following two models: (i) the price gyration algorithm for searching initial values and the LMA for optimisation, M1, and (ii) the improved price gyration algorithm, which is applied with floating window sizes and a distance-based weighting approach, and a constrained GA for optimisation, M2.

For both methods, we forecast the critical time  $t_c$  using different final observations of the estimation sample from one to three months before the actual critical time and merge the results into one sample. The probability density functions of estimated critical times  $t_c$  are shown in Figure 4. We can observe that M1 does not predict critical time well because it produces  $t_c$  either far away from the actual critical time such as NASDAQ or strongly biased such as NIKKEI, HSI, and SSEC. Put differently, M2 predicts a relatively accurate  $t_c$  compared with M1. Most of the predicted critical times  $t_c$  are close to each other and highly concentrated around the actual critical time, which demonstrates that our method, M2, outperforms M1. Detailed comparison with other estimation algorithms employed in the literature is presented in Appendices A and B.

## 5. Conclusions

This paper provides an algorithm to predict the critical time of financial bubbles with an LPPL model. The parameters are estimated by minimising the cost function by means of a nonlinear optimisation method. This algorithm consists of two steps: (i) a price gyration method to generate an initial candidate of parameters and (ii) a GA to find the optimal solution. Specifically, we go beyond the price gyration method in the previous literature. In our case, different window sizes are applied to peak detection since the fixed window size may omit the possible variation in the cycle of LPPL growth. Given the peaks detected, we assign the distance-based weights on each peak according to its distance from the last observation in the estimation sample. The distance-based weighting method makes the estimation accord with reality; that is, the recent data include more information on forecasting. We also use simulated annealing to optimise the model and find that the results are almost the same as those of a GA (up to  $10^{-4}$ ). Therefore, we attribute the success of our algorithm to the price gyration method with different window sizes and the distance-based weighting approach.

For validation, we perform an ex-ante prediction on the time of crashes on four stock market indices. The critical time of the bubbles, when the crashes may happen with significant probability, is one of the parameters in the LPPL model. Our predictions on critical time are highly concentrated around the actual time. Moreover, a diagnostic analysis demonstrates our results in different aspects. First, we generate smaller prediction errors for the LPPL model. Second, our prediction is stable with respect to varying the termination time of the observation period. Third, in terms of the degree of concentration and accuracy, we present a more significant and robust improvement than existing algorithms.

Our work focuses on bubbles and crashes in stock prices, and it can be further extended to other assets, such as real estates and foreign exchange rates, where both bubbles and herding behaviour can be observed, because the LPPL model describes the dynamics of financial bubbles generated by positive feedback mechanisms, especially the herding behaviour of noise traders [19–22]. However, the LPPL model mainly focuses on the time period prior to crashes, therefore it leaves blank the price behaviour in the postcrash periods even when mispricing may still exist, which is of policy significance in order to stabilise and boost the economy once crashes occur. In addition, since the parameters in the LPPL characterise stock markets, they have a considerable potential to be related to economic fundamentals. Thus, as future research, additional analysis of whether and how the LPPL model reflects macroeconomic conditions would be of interest.

## Appendix

### A. Computational Efficiency and Accuracy

We evaluate the computational efficiency of the four algorithms in terms of running time per each iteration. The whole computational effort per each iteration is the sum of running time needed to generate a set of initial values and to minimise the RMSE. We compare the computational efficiency of our algorithm with that of others in Table 6 using the NASDAQ case. The number of necessary iterations  $I$  and prediction results, that is, either (i) point estimate and its bias against actual crash date ( $\hat{t}_c - t_c$ ) or (ii) interval estimate (95% prediction interval) and  $P_{60}$ , are also reported.

The *Tabu search* algorithm for generating initial values and the LMA for optimisation, named as TL, and GA provide a point estimator about  $t_c$ , which is the only statistic that these algorithms yield by minimising the RMSE. However, they have no way of conducting a hypothesis test and justifying their clinical usage. The other algorithm, M1, runs even more efficiently than the first two and is able to construct a confidence interval. However, it generates estimation results on a broad spectrum and leads to a wide prediction interval and poor values of  $P_{60}$ . On the contrary, our algorithm, M2, achieves precise estimation with a narrow prediction interval, and most estimates locate within 60 weekdays of the actual crash date. Since our algorithm includes GA, it takes as much time as other global optimisation approaches, such as *Tabu search* and pure GA. Especially when a large number of iterations is implemented, the efficiency of our algorithm is much lower than that of others. We, however, can exploit parallel computation on a general-purpose multicore process to heavily cut down the computational time. In short, there is a trade-off between estimation accuracy and computational efficiency.

### B. Estimating Algorithms

We extensively review the literature and conclude that there are five different categories of algorithms for LPPL estimation: (i) *2-step nonlinear optimisation* [4]; (ii) GA [12, 14];

TABLE 6: Computational effort of the four different algorithms.

Algorithm	Computational effort	$I$	Prediction results
TL	$10 + 10^{-1} \approx 10s$	10	04/02/2000 $\hat{t}_c - t_c = -24$
GA	10s	10	19/05/2000 $\hat{t}_c - t_c = 49$
M1	$10^{-4} + 10^{-1} \approx 10^{-1}s$	2500	11/01/2000–13/06/2012 $P_{60} = 23\%$
M2	$10^{-2} + 10 \approx 10s$	2500	13/01/2000–17/05/2000 $P_{60} = 99\%$

(iii) *2-step/3-step maximum likelihood approach* [18]; (iv) *parallel LMA* [8]; and (v) *modified profile likelihood* [23]. We further provide a comparison between three major approaches and ours in Appendix A.

One of the most common approaches is the *2-step nonlinear optimisation*. First, 10 potential sets of solutions can be obtained using a so-called *Tabu search* [24]. Second, each of these 10 solutions is used as a set of initial values in an LMA, a nonlinear least square algorithm. Finally, the set of parameters that produces the minimum sum of squared errors between the fitted model and the observations is taken as the final solution [4].

As an alternative to relying on gradient-based optimisation methods, such as LMA, a GA can be utilized for LPPL estimation. GA is a metaheuristic algorithm motivated by the process of natural selection belonging to the larger class of evolutionary algorithms. GA is commonly employed to obtain high-quality solutions of optimisation including searching problems by relying on bio-inspired operators such as mutation, crossover, and selection. GA overcomes the problem of local minima and allows for the setting of reasonable constraints on parameter spaces [12, 14].

Unlike the two methods above which find the best fit by minimising the sum of squared errors (as a cost function), Fantazzini [18] proposed a *2- or 3-step maximum likelihood* (ML) for a generalised LPPL model in which strong autocorrelation and heteroscedasticity of residual terms are allowed. This approach estimates the LPPL model with a reversed time series called “anti-bubbles.” The *2-step approach* estimates the original LPPL parameters and other parameters related to residual terms separately in two stages. The *3-step approach* includes one more step at the beginning of the *2-step approach* to provide the set of initial values for LPPL parameters. This ML optimisation method employed the Broyden, Fletcher, Goldfarb, Shanno (BFGS) algorithm that searches for a zero gradient. Similar to other gradient-based optimisation approaches, the main disadvantage of this approach is that it can be easily trapped in local minima.

One of the weaknesses of those algorithms is that they can only provide point estimation results, and thus the statistical robustness of the results is low. Liberatore [8] proposed a price gyration method to provide initial solutions to LMA based on the peaks of price trajectory. Compared to other methods such as *Tabu search* that provides a set of initial values, the price gyration method utilizes richer

information, presented in the microstructure of price trajectory. With the abundant initial values generated by price gyration, we can obtain a distribution of LPPL parameters for interval estimation.

A recent paper [23] used a *modified profile likelihood inference* to provide interval estimation for the critical time in the LPPL model. It takes into account the impact of other nonlinear (or “nuisance”) parameters for correct adjustment of the uncertainty regarding the critical time. Similar to *modified profile likelihood*, our method provides interval estimation. However, we incorporate parameter uncertainty not by assuming a certain probability distribution of error terms but rather through price gyration, which systematically generates a set of initial values in the parameter spaces.

Our algorithm, M2, is closely related to *parallel LMA*, M1. The first step of price gyration is to detect peaks, which are a local maximum within a window. Once the peaks are detected, price gyration might encounter certain problems. Specifically, the prediction results can be unstable for different window sizes, and the estimation of the critical time is not sufficiently accurate if the peaks are too far from the last observation of the sample. Hence, we further improve the price gyration method in two ways. First, we detect peaks by varying the window size, which allows for the possibility of different lengths of cycle. Second, we implement a distance-based weighting approach for peak selection; that is, the more recent the peak is, the greater the probability it will be selected. In addition, we substitute GA for LMA to search for the global minima. Furthermore, we also compare M2 with the original *2-step nonlinear optimisation* and pure GA, both of which only provide point estimation, presented in Table 6.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Disclosure

This paper was formerly circulated under the title “Predicting the Critical Time of Financial Bubbles.”

## Conflicts of Interest

The authors of this research declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by KAIST through the 4th Industrial Revolution Project and Moon Soul Graduate School of Future Strategy (Kwangwon Ahn), and by Peking University HSBC Business School through Bairui Trust Research Fund (Kwangwon Ahn). In addition, the authors would like to thank Kyungmoo Heo, Hanwool Jang, and Guseon Ji for their excellent research assistant work.

## References

- [1] J. E. Stiglitz, "Symposium on bubbles," *Journal of Economic Perspectives*, vol. 4, no. 2, pp. 13–18, 1990.
- [2] F. Allen, S. Morris, and A. Postlewaite, "Finite bubbles with short sale constraints and asymmetric information," *Journal of Economic Theory*, vol. 61, no. 2, pp. 206–229, 1993.
- [3] M. K. Brunnermeier, "Bubbles," in *Banking Crises*, G. Jones, Ed., Palgrave Macmillan, London, 2016.
- [4] A. Johansen, O. Ledoit, and D. Sornette, "Crashes as critical points," *International Journal of Theoretical and Applied Finance*, vol. 3, no. 2, pp. 219–255, 2000.
- [5] A. Clark, "Evidence of log-periodicity in corporate bond spreads," *Physica A: Statistical Mechanics and its Applications*, vol. 338, no. 3–4, pp. 585–595, 2004.
- [6] V. Filimonov and D. Sornette, "A stable and robust calibration scheme of the log-periodic power law model," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 17, pp. 3698–3707, 2013.
- [7] A. Johansen and D. Sornette, "Bubbles and anti-bubbles in Latin-American, Asian and Western stock markets: an empirical study," *International Journal of Theoretical and Applied Finance*, vol. 4, no. 6, pp. 853–920, 2001.
- [8] V. Liberatore, "Computational LPPL fit to financial bubbles," *arXiv preprint*, pp. 1003–2920, 2011.
- [9] D. T. Pele, "An LPPL algorithm for estimating the critical time of a stock market bubble," *Journal of Social and Economic Statistics*, vol. 1, no. 2, pp. 14–22, 2012.
- [10] J. R. Kurz-Kim, "Early warning indicator for financial crashes using the log periodic power law," *Applied Economics Letters*, vol. 19, no. 15, pp. 1465–1469, 2012.
- [11] P. Geraskin and D. Fantazzini, "Everything you always wanted to know about log-periodic power laws for bubble modeling but were afraid to ask," *The European Journal of Finance*, vol. 19, no. 5, pp. 366–391, 2013.
- [12] P. Korzeniowski and I. Kuropka, "Forecasting the critical points of stock markets indices using log-periodic power law," *Ekonometria*, vol. 1, pp. 100–110, 2013.
- [13] G. Palshikar, "Simple algorithms for peak detection in time-series," in *Proceedings of the 1st IIMA international conference on advanced data analysis, business analytics and Intelligence*, pp. 1–13, Ahmedabad, India, 2009.
- [14] E. Jacobsson, "How to predict crashes in financial markets with the log-periodic power law," *Master Dissertation, Department of Mathematical Statistics*, Stockholm University, 2009.
- [15] L. Lin, R. E. Ren, and D. Sornette, "The volatility-confined LPPL model: a consistent model of 'explosive' financial bubbles with mean-reverting residuals," *International Review of Financial Analysis*, vol. 33, pp. 210–225, 2014.
- [16] A. Johansen, "Characterization of large price variations in financial markets," *Physica A: Statistical Mechanics and its Applications*, vol. 324, no. 1–2, pp. 157–166, 2003.
- [17] D. S. Brée and N. L. Joseph, "Testing for financial crashes using the log-periodic power law model," *International Review of Financial Analysis*, vol. 30, pp. 287–297, 2013.
- [18] D. Fantazzini, "Modelling bubbles and anti-bubbles in bear markets: a medium-term trading analysis," in *Handbook of Trading: Strategies for Navigating and Profiting from Currency, Bond, and Stock Markets*, G. N. Gregoriou, Ed., pp. 365–388, McGraw-Hill Education, New York, 2010.
- [19] W. T. Woo, "Some evidence of speculative bubbles in the foreign exchange markets," *Journal of Money, Credit and Banking*, vol. 19, no. 4, pp. 499–514, 1987.
- [20] W.-X. Zhou and D. Sornette, "Is there a real-estate bubble in the US?," *Physica A: Statistical Mechanics and its Applications*, vol. 361, no. 1, pp. 297–308, 2006.
- [21] P. R. Kaltwasser, "Uncertainty about fundamentals and herding behavior in the FOREX market," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 6, pp. 1215–1222, 2010.
- [22] V. Babalos, M. Balcilar, and R. Gupta, "Herding behavior in real estate markets: novel evidence from a Markov-switching model," *Journal of Behavioral and Experimental Finance*, vol. 8, pp. 40–43, 2015.
- [23] V. Filimonov, G. Demos, and D. Sornette, "Modified profile likelihood inference and interval forecast of the burst of financial bubbles," *Quantitative Finance*, vol. 17, no. 8, pp. 1167–1186, 2017.
- [24] D. Cvijovic' and J. Klinowski, "Taboo search: an approach to the multiple minima problem," *Science*, vol. 267, no. 5198, pp. 664–666, 1995.

