

자산 가격 버블의 붕괴 위험 평가를 위한 딥러닝 모델

김진우, 안광원

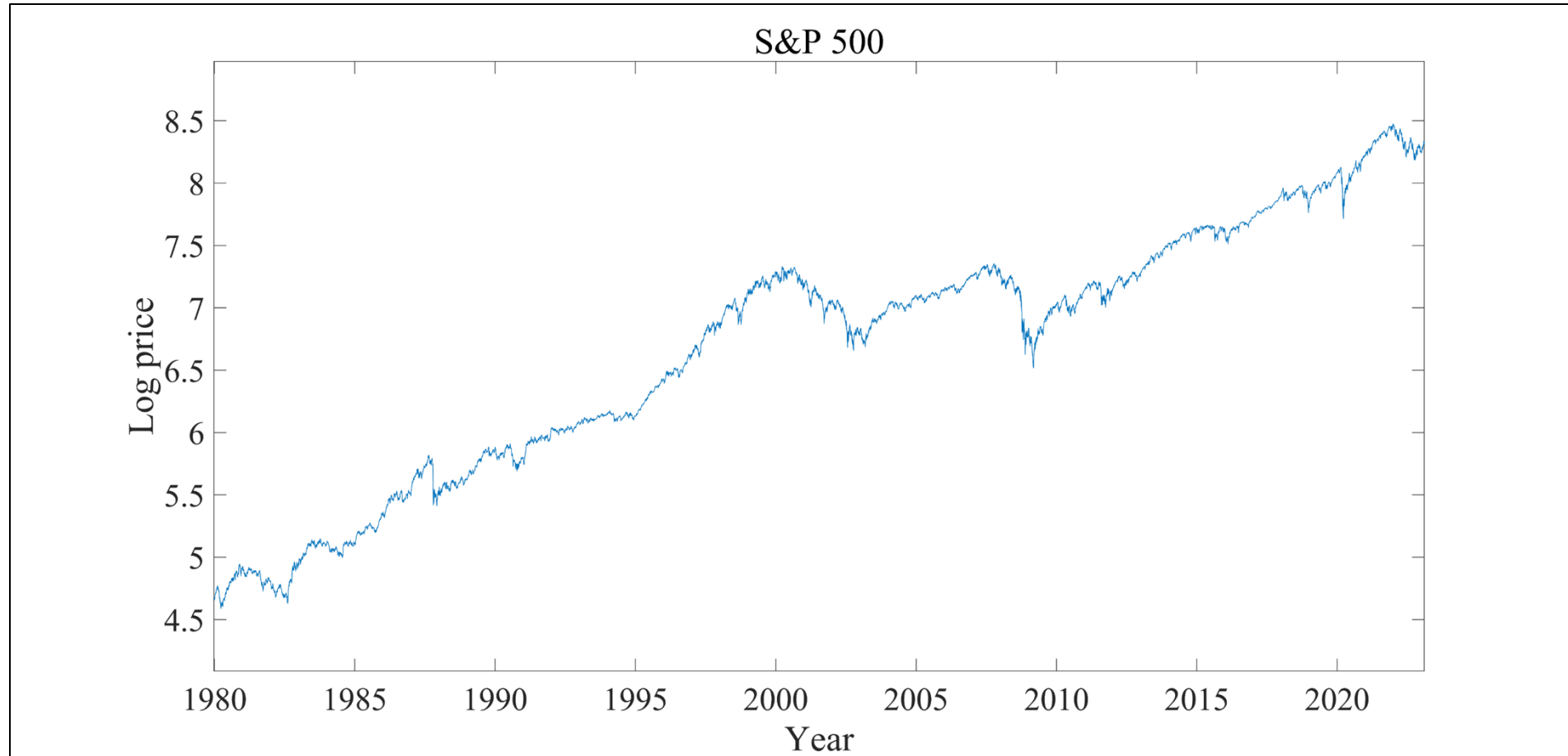
연세대학교 산업공학과, 금융기술센터

jinu.kim@yonsei.ac.kr

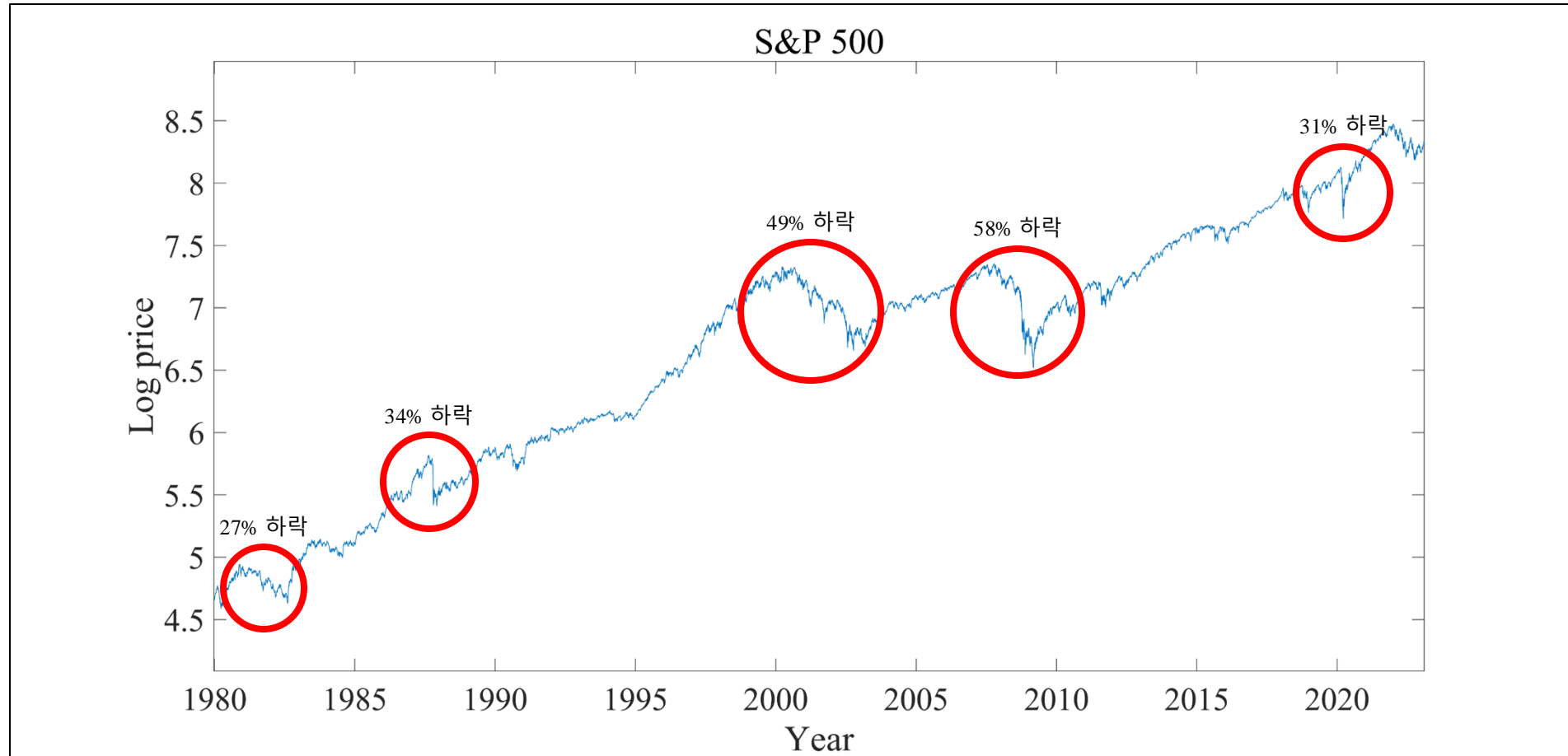
Outline

1. Introduction
2. Data and Methodology
3. Results and Discussion
4. Conclusion

1. Introduction

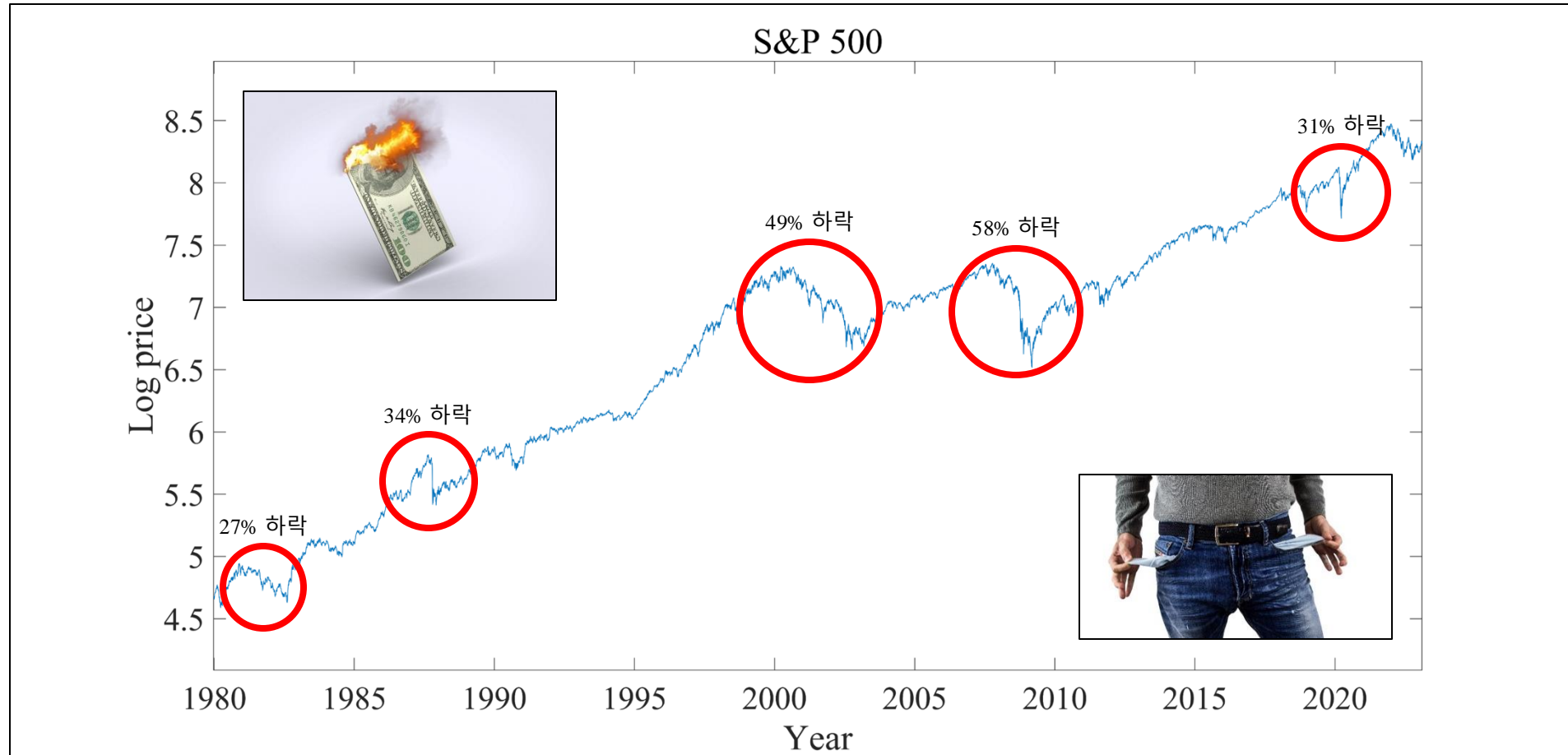


1. Introduction



역사적으로 금융 시장의 붕괴는 주식 가격 폭락과 함께 경제 위기 등을 동반하여 많은 사람에게 큰 손실을 유발하였다.

1. Introduction



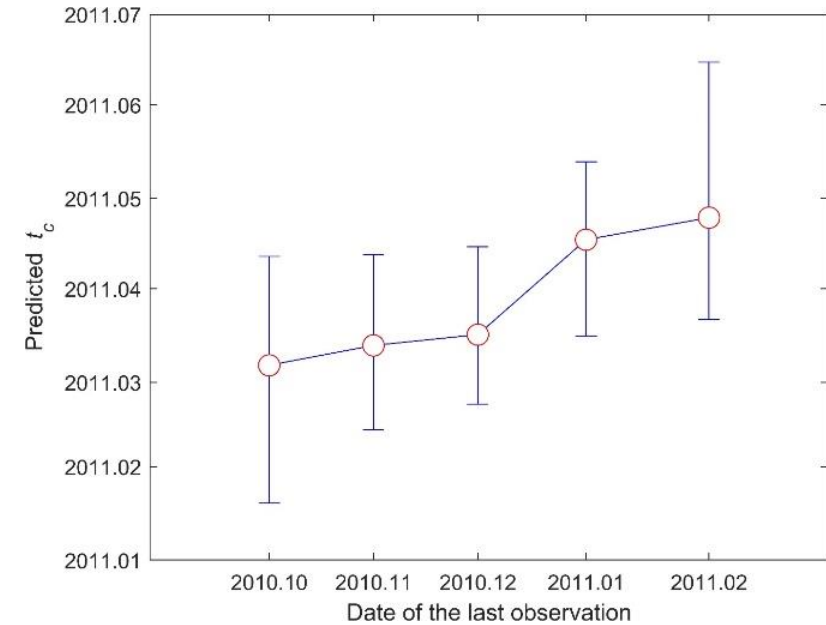
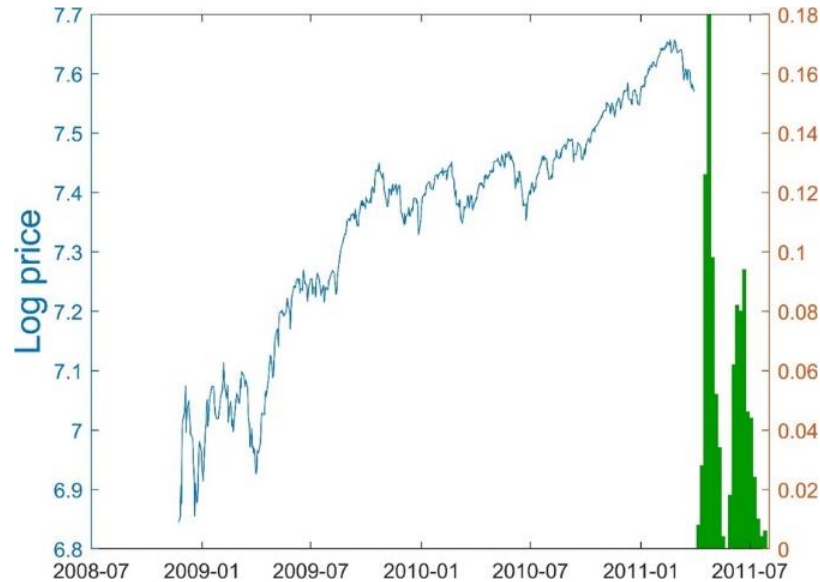
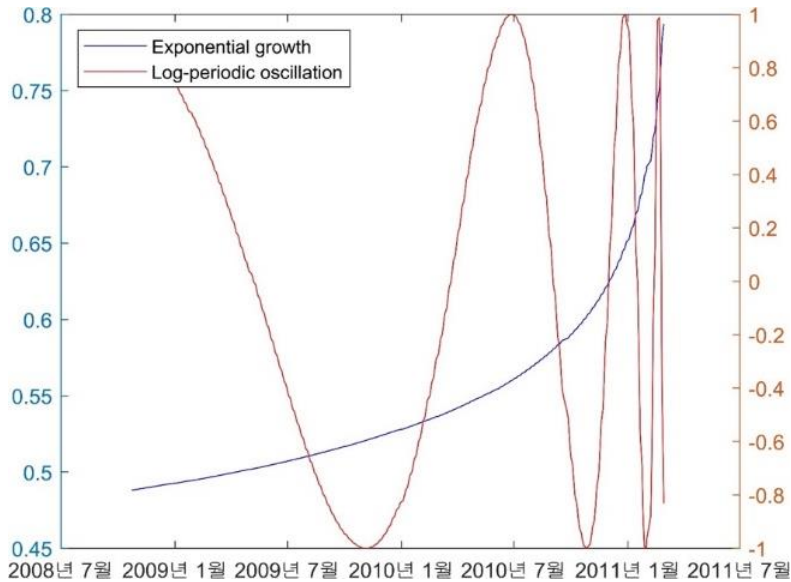
버블 붕괴 = 손실?

1. Introduction

- 자산 가격의 붕괴를 예측하기 위한 연구가 진행되었고, 그 중 log-periodic power law (LPPL) 모델은 자산 가격 버블의 붕괴 전에 나타나는 전조 현상에 근거해서 그 붕괴 시점을 추정한다.
- Johansen et al. (2000)이 지진을 예측하기 위해 고안한 모델을 발전시켜 제안했으며, 많은 선행 연구에서 LPPL 모델을 사용하여 과거의 자산 가격 버블 붕괴를 효과적으로 예측하였다 (Dai et al., 2018; Ahn et al., 2023).
- 본 연구에서는 LPPL 모델과 딥러닝 모델을 결합하여, LPPL 모델의 추정 결과를 보다 실용적으로 활용할 수 있게 발전시켰다.

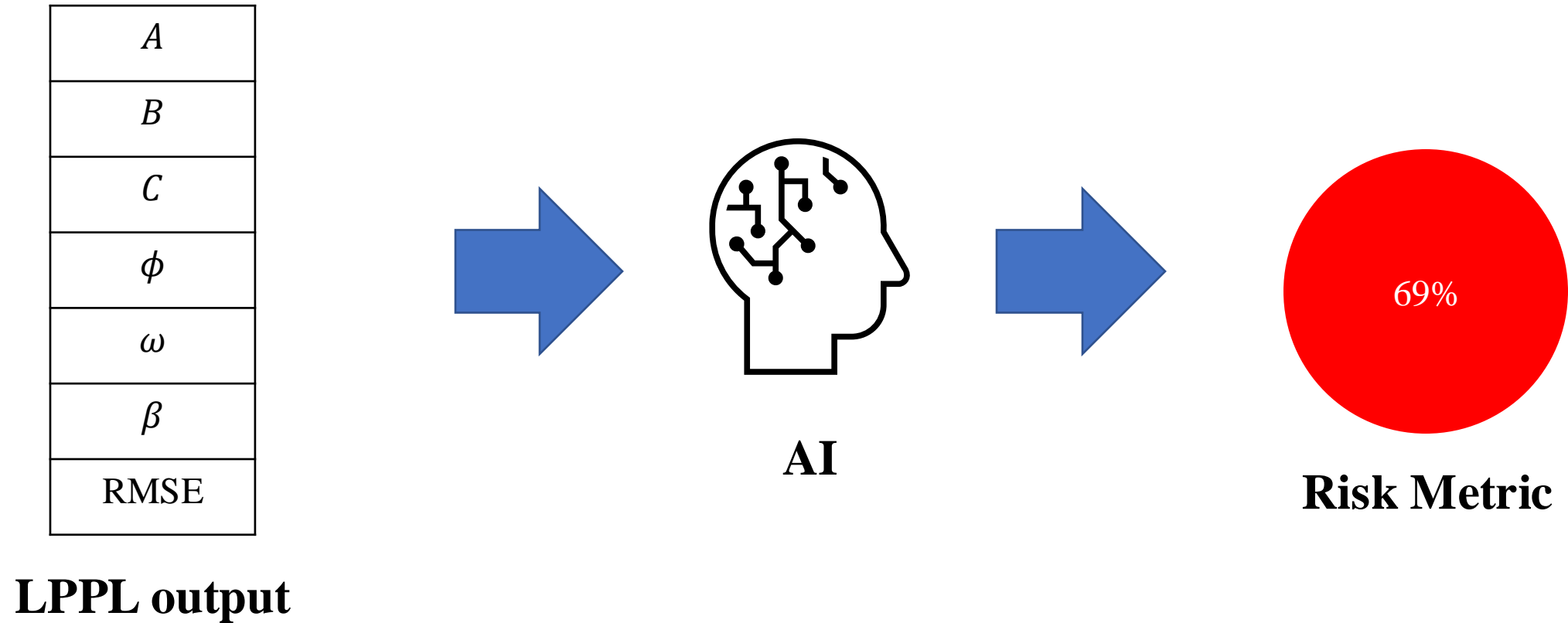
1. Introduction

A	B	C	t_c	ϕ	ω	β
-----	-----	-----	-------	--------	----------	---------



그러나 LPPL 추정 결과의 해석과 신뢰성 판단은 연구자의 주관적 판단과 역량에 의존하였다.

1. Introduction



따라서 본 연구는 인공지능 모델을 활용하여 LPPL 추정 파라미터(t_c)의 신뢰도를 계량화하고자 한다.
나아가 자산 가격 붕괴 위험의 척도로 사용할 수 있는 위험 지표를 개발하고자 한다.

2. Data and Methodology

■ 데이터

- 자산 가격 데이터

- 1990년 이후 상장된 주식 중 S&P500의 시가총액 상위 100개 종목의 일별 종가 데이터
(출처: yahoo finance)

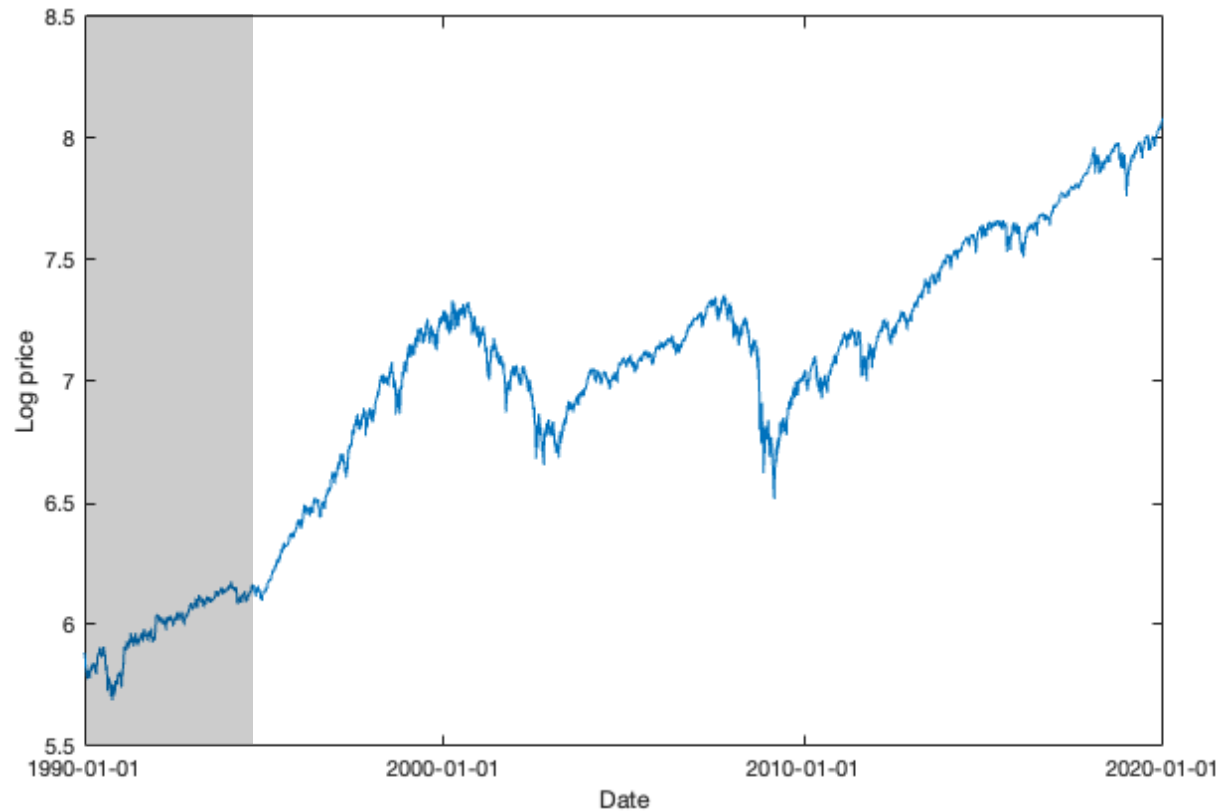
- S&P 500은 세계에서 가장 크고 영향력 있는 지수로, 세계 증시를 대표할 수 있는 일반적으로 사용되는 벤치마크 지수이기 때문에 분석 대상으로 선정하였다.

- 수집 데이터 기간: 1990.01—2019.12 (30y)

2. Data and Methodology

■ 데이터

- 인공지능 데이터
 - 데이터 기간: 1995.01—2019.12 (25y)



2. Data and Methodology

■ 데이터

- 인공지능 데이터

Time	A	B	C	t_c	ϕ	ω	β	RMSE	Label
1995.01.02	-1.04	-0.01	-0.83	36.37	0.01	6.00	0.37	0.01	0
1995.01.03	4.50	-0.11	0.06	178.78	4.20	6.64	0.36	0.02	1

- 레이블링을 진행하는 데 있어서, 추정 t_c 의 ± 5 trading day의 추정 오차를 허용
- 하나의 Time에 대해서 500개의 LPPL 파라미터를 추정

2. Data and Methodology

Table 1. The number of data points

	Crash	Non-Crash	Total
Log return	1,145 (0.18%)	628,155 (99.82%)	629,300
	Label 1	Label 0	Total
LPPL estimation	222,403 (1.61%)	13,627,579 (98.39%)	13,849,982

2. Data and Methodology

■ LPPL (Johansen et al., 2000)

- Solution to the price dynamics:

$$dp_t = u_t p_t dt - \kappa p_t dj$$

$$\log \frac{p_t}{p_{t_0}} = \kappa \int_{t_0}^t h_s ds$$

u_t : drift

j : jump process

κ : jump size (the drop ratio after crash)

h_t : the hazard rate

- Hazard Rate:

$$h_t \approx B_1 (t_c - t)^{-(1-\beta)} \{1 + C_1 \cos[\omega \log(t_c - t) + \varphi]\}$$

t_c : the critical time

2. Data and Methodology

■ LPPL (Johansen et al., 2000)

- Log-periodic power law:

$$Y_t \approx A + B \cdot (t_c - t)^\beta \{1 + C \cos[\omega \log(t_c - t) + \phi]\}$$

■ Fitting the LPPL Parameters

- Optimization: Genetic algorithm
- Evaluation metric: Root mean squared error

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

2. Data and Methodology

Table 2. LPPL parameters

Parameter	Constraint	Meaning	
A	$(\max P, +\infty)$	(1) The price at critical time; (2) Higher than the current price	
B	$(-\infty, 0)$	Negative, measuring the distance to critical price	
C	$C \in (-1, 1)$	Oscillations cannot overwhelm the trend	
t_c	$(t, +\infty)$	(1) The most probable critical time; (2) Haven't crashed	
ϕ	$[0, 2\pi)$	Phase adjustment	
		The frequency of the fluctuations during the bubble	
ω	$[4.8, 13]$	Too small	False oscillation
		Too large	Fit the random noises
		The exponent of the power law growth	
β	$[0.1, 0.9]$	Negative β	Infinite critical price
		Positive but close to 0	No trend
		Positive but close to 1	Low hazard rate

2. Data and Methodology

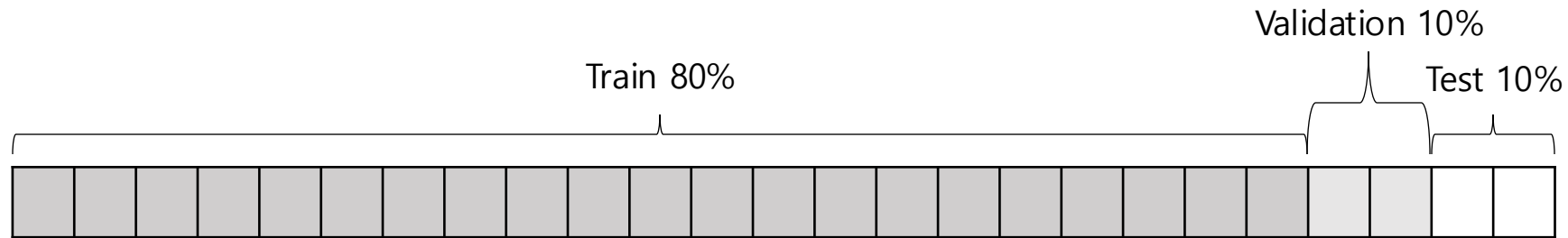
■ DNN based classification model

- 입력 변수는 파라미터 A , t_c 를 제외한 6개의 변수($B, C, \phi, \omega, \beta, RMSE$)를 사용하였다.
- 제안하는 모델은 총 3개의 은닉층으로 구성되어 있으며, 각 은닉층은 128, 64, 32개의 노드로 이루어진 신경망이다.
- 데이터 불균형 문제를 해결하기 위해 SMOTE(Synthetic minority oversampling technique)을 사용하였다.

2. Data and Methodology

■ DNN based classification model

- Train/Validation/Test period



- Evaluation metric
 - Accuracy = $\frac{TP + TN}{TP + TN + FP + FN}$
 - Precision = $\frac{TP}{TP + FP}$
 - Recall = $\frac{TP}{TP + FN}$
 - F1 score = $\frac{2 * TP}{2 * TP + FP + FN}$

3. Results and Discussion

■ DNN based classification model

Model	Input variables	Accuracy	Recall	Precision	F1 score
AI (DNN)	$B, C, t_c, \phi, \omega, \beta$	56.35%	50.09%	57.26%	53.44%
Non-linear ML (Random forest)	$B, C, t_c, \phi, \omega, \beta$	49.72%	1.58%	42.61%	3.05%
Linear ML (Logistic regression)	$B, C, t_c, \phi, \omega, \beta$	50.71%	41.20%	50.88%	45.53%

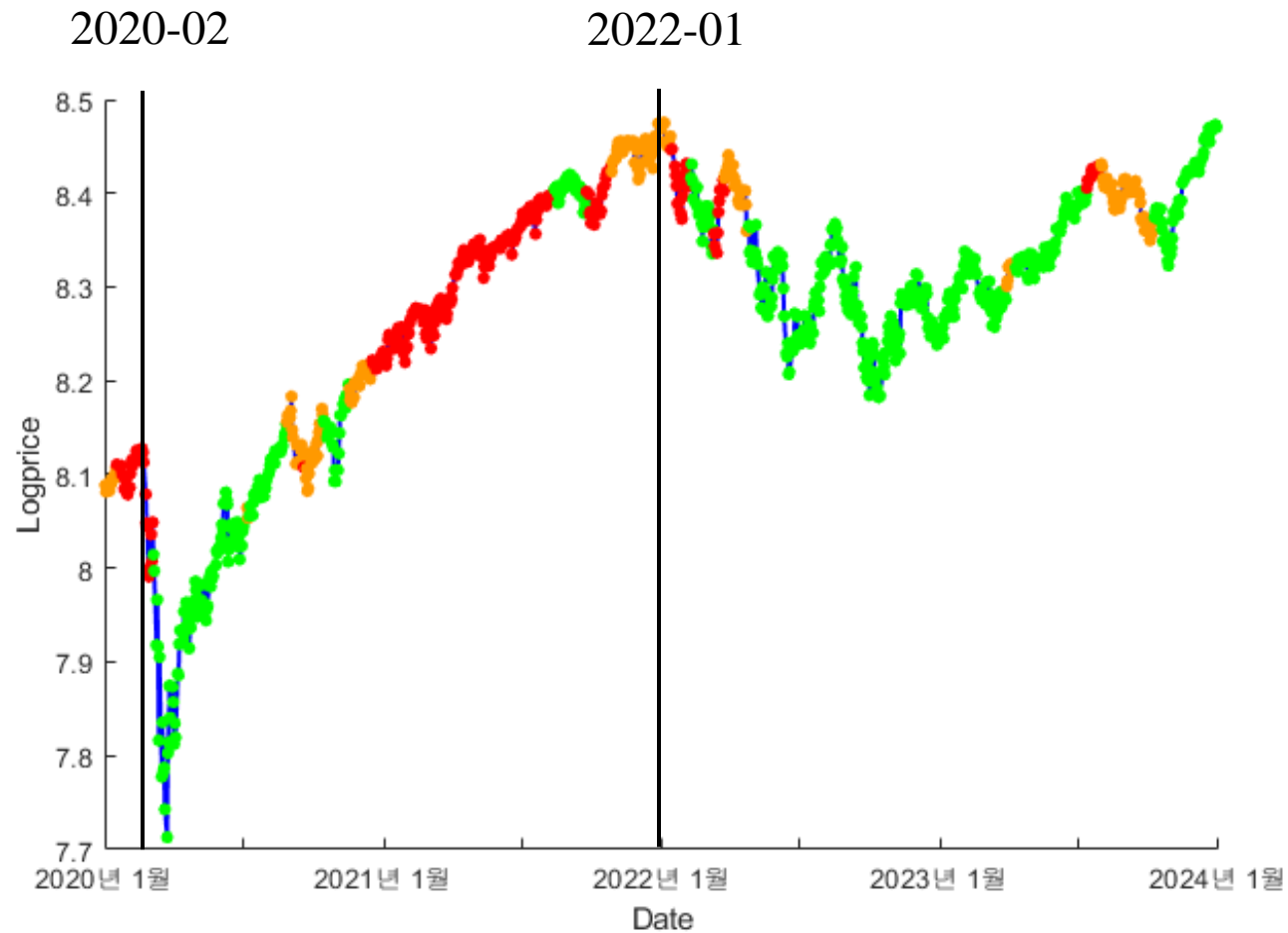
3. Results and Discussion

■ DNN based classification model

- Test data에 대해 정확도, 재현율, 정밀도, F1 score 모든 지표에서 제안하는 DNN 모델이 선형 ML, 비선형 ML 보다 더 높은 성과를 보인다. 이는 LPPL 추정 파라미터의 신뢰도 판단 문제가 복잡한 비선형 패턴을 보이기 때문으로 판단된다.
- 드물게 발생하는 자산 가격의 붕괴는 큰 손실을 가져오기 때문에 보수적인 자세로 위험을 최소화하는 것이 중요하다. 따라서 57.26%의 정밀도는 43.73%의 오경보를 감안하더라도 자산 가격의 붕괴에 의한 손실을 최소화하고 이익을 창출하기에 충분한 수치이다.
- 나아가 56.35% 정확도와 전체 붕괴 중 50% 이상을 성공적으로 예측하는 재현율은 해당 모델이 유의미한 정보를 금융 시장 참여자에게 제공할 수 있음을 시사한다.

3. Results and Discussion

■ Application: Risk metric



3. Results and Discussion

■ Application: Risk metric

- 인공지능 모델 구축에 사용된 데이터의 이후 기간인 2020년부터 2023년 까지의 S&P 500 지수 데이터에 대한 위험 지표는 2020년 2월과 2022년 1월 발생한 자산 가격 버블의 붕괴에 대한 위험을 사전에 비교적 효과적으로 감지한다.
- 해당 결과를 통해 본 연구에서 제안하는 위험 지표가 자산 가격 버블의 붕괴 위험을 적절히 계량화하는 것으로 판단할 수 있다.

4. Conclusion

Summary

- 본 연구는 LPPL 추정 파라미터의 신뢰도를 판단하는 딥러닝 모델을 구축하고, 이를 활용한 새로운 위험 지표를 제안하였다.
- 제안하는 모델은 벤치마크 모델에 비해 추정 파라미터의 신뢰도 판단에 있어서 좋은 성능을 보여주었으며, 개발한 위험 지표는 위험을 적절히 감지하였다.

4. Conclusion

Future work

- 향후 연구에서는 시계열 데이터의 특성을 감안하여 시계열 데이터에 주로 사용하는 인공지능 모델을 활용하거나 학습 데이터의 window size를 조절함으로써 분류 성능을 높이는 시도를 할 것이다.
- Risk metric의 평가 지표를 설정하고, 이를 최적화하는 방향으로 발전시켜 나가는 연구를 진행할 것이다.

4. Conclusion

Contribution

- 본 연구는 기존 LPPL에 대한 연구들과 달리 파라미터 추정 방법이나 실증 분석이 아닌 추정된 파라미터를 어떻게 해석하고 실용적으로 활용할 것인가에 대해 연구하였다.
- 본 연구의 결과를 통해 투자자와 정책입안자는 자산 가격 버블의 붕괴 위험을 사전에 분석하고 효과적으로 대응할 수 있을 것으로 기대한다.