봉수 예측지수(BPI)와 실적배수의 실증 설계 — 봉인 11 건의 검증과 운·실력 구분에 필요한 채점 건수

채점 0 건에서 출발하는 예측지수: 무엇을 봉인했고, 몇 건이 쌓여야 실력이라 말할 수 있나



핵심 주제

예측지수 · 봉인(SHA-256) · 브라이어 점수 · 기술점수 · 실적배수 · 예측 성과 보상 · 몬테카를로 모의

저자: 조봉수(봉수 SSGlobal(주) 대표이사, 필명). 투자 권유 아님.

정식본 2026 년 10 월

목차

국문 요약 3
1. 서론 **4**
1.1 문제 제기 4
1.2 연구 질문 4
2. 지수 정의 **4**
2.1 봉인과 공개 4
2.2 채점 산식 4
3. 실증 **1** — 봉인 장부의 무결성**(**실측**)** **5**
4. 실증 **2** — **11** 건 채점의 운과 실력**(**모의**)** **5**
5. 실증 **3** — 몇 건이 쌓여야 하나**(**모의**)** **6**
5.1 가격 경로(보정된 예측을 가정한 모의 중앙값) 7
6. 논의와 정책 제안 **7**
7. 한계 **8**
8. 결론 **8**
참고문헌 **8**
부록 **A.** 인공지능 이용 고지 **9**
부록 **B.** 재현 방법 **9**

국문 요약

봉수 예측지수(Bongsu Prediction Index, BPI)는 채점 전에 봉인·공개한 확률 예측 만을 채점하는 지수다. 브라이어 점수로 '반반 찍기' 대비 기술점수(BSS)를 구하고, 채점 건수에 따른 신뢰계수 n/(n+20)을 곱한다. v1.5 정책에서는 같은 기술점수로 실적배수 1 + max(0, BSS) × n ÷ 5 를 계산해 예측이 들어간 상품의 가격에 곱한다. 이 배수에는 상한이 없고, 9,900 원을 넘는 부분은 예측 성과 보상 이 된다.

2026 년 10 월 5 일 현재 채점이 끝난 예측은 **0** 건 이다. 따라서 BPI 는 0, 실적배수는 1.0 이다. 이 논문은 실적이 없다는 사실을 출발점으로 삼아 두 가지를 실증한다.

① 봉인 장부의 무결성**(**실측**).** 논술 분야 봉인 예측 11 건의 SHA-256 해시를 다시 계산하였다. 결과는 다음과 같다.

- 11 건 모두 원장과 일치했다.
- 11 건 모두 같은 날 공개한 해시 목록에 들어 있었다.
- 산출이력 1 줄도 해시 체인이 정상이었다.

예측 확률은 0.42~0.90(평균 0.697)이고, 기준확률은 모두 0.50 이다. 채점일은 2026 년 10 월 11 일부터 12 월 6 일까지다.

② 운과 실력의 구분**(**모의**,** 실제 결과 아님**).** 이 11 건의 확률 그대로 20,000 번 모의 채점하였다. 결과를 세 가지 세계로 나누어 보면 다음과 같다.

가정한 세계	기술점수 **BSS** 가 **0** 보다 클 확률	실적배수 중앙값
확률대로 일어나는 세계(보정된 예측)	86%	1.6 배
실력이 없는 세계(모두 50%)	22%	1.0 배
과신한 세계	52%	—

실력이 없는 세계에서도 배수가 우연히 1.0 을 넘을 확률은 22%였고, 95 분위로는 1.57 배까지 갈 수 있었다. 채점이 55 건 쌓이면 이 우연 확률은 3%, 99 건이면 1% 아래로 떨어진다. 같은 확률 수준을 유지한 보정된 예측이 C 등급 상품(2,200 원)을 9,900 원 위로 올리는 데는 약 **99** 건의 채점 이 필요하였다(모의 중앙값 5.9 배, 13,000 원).

결론적으로 지금의 실적배수 식은 적은 건수에서 운으로 배수를 받을 여지가 있다(11 건에서 최대 약 1.6 배). 대신 실력이 있으면 수십~수백 건에서 배수가 크게 벌어진다. 봉수는 이 한계를 공개하고, 첫 채점(2026-10-12 가톨릭대) 결과를 맞은 것과 틀린 것 모두 그대로 공개한다.

주제어: 예측지수, 봉인, 브라이어 점수, 기술점수, 실적배수, 성과 보상, 몬테카를로

1. 서론

1.1 문제 제기

정보상품 시장에서 “적중했다”는 주장은 흔하지만, 언제 무엇을 몇 %로 말했는지 미리 남기지 않으면 검증할 수 없다. 맞힌 것만 골라 말하는 선택 보고가 가능하기 때문이다. 봉수는 이를 막으려고 예측을 시험·사건 전에 **SHA-256** 지문으로 봉인하고 그 지문을 먼저 공개 한 뒤, 채점일이 지나면 맞은 것과 틀린 것을 모두 공개한다(봉수의 원칙 “봉인하고, 채점하고, 공개한다”).

v1.5 정책은 여기에 경제적 의미를 더했다. 검증된 실적이 쌓이면 성장 기업의 주가수익비율(PER)처럼 가격 배수가 커지고, 상한이 없다. 따라서 배수가 운으로 오르지 않는지 를 미리 따져 보는 것이 공정한 성과 보상의 전제다.

1.2 연구 질문

- 연구 질문 1: 지금 봉인된 예측은 변조 없이 공개되었는가?
- 연구 질문 2: 11 건만 채점되었을 때 기술점수와 배수는 어디까지 흔들릴 수 있는가?
- 연구 질문 3: 운과 실력을 가르려면 몇 건이 채점되어야 하는가?
- 연구 질문 4: 실력이 있다면 가격은 어떤 경로로 오르는가?

2. 지수 정의

2.1 봉인과 공개

봉인 문자열은 예측 ID·상품 ID·봉인일·채점예정일·예측내용·확률·기준확률을 '|'로 이은 것이며, 그 SHA-256 지문을 원장과 공개 해시 목록에 함께 남긴다. 문구나 확률을 한 글자라도 고치면 지문이 맞지 않아 지수에서 자동으로 빠진다. 채점은 결과(0/1)와 함께 판정 이유·근거 **URL** 을 채점근거기록에 남긴 것만 센다.

2.2 채점 산식

- 브라이어 점수 BS = (1/n) Σ (p − o)²
- 기준 브라이어 BS_ref = (1/n) Σ (b − o)², 기준확률 b(근거 없으면 0.5)
- 기술점수 BSS = 1 − BS ÷ BS_ref
- 신뢰계수 Z = n ÷ (n + 20)
- **BPI = 100 × max(0, BSS) × Z** (0~100)
- 실적배수 **= 1 + max(0, BSS) × n ÷ 5** (최저 1.0, 상한 없음)[1]

[1] 실적배수는 2026-10-05 대표 결정(공지 v1.5)이다. 예측이 들어간 상품에만, 분야별로 적용한다. 기술점수가 0 이하이면 1.0 배로 돌아간다.

브라이어 점수는 확률 예측의 표준 채점법이다.[2] 기준확률보다 잘 맞혀야 BSS 가 양수가 되므로, 쉬운 것을 99%라고 말해 점수를 부풀리는 일을 막는다.

3. 실증 1 — 봉인 장부의 무결성(실측)

표 1. 봉인 장부 점검 결과(2026-10-05)

항목	결과
봉인 예측 수(논술 분야)	11 건
해시 재계산 일치	11/11
공개 해시 목록 포함	11/11
근거가 기록된 채점	0 건
현재 BPI · 실적배수	0 · 1.0 배
예측 확률	0.42~0.90 (평균 0.697)
기준확률	모두 0.50
채점예정일	2026-10-11, 10-18, 11-22, 11-27, 12-05, 12-06
산출이력	1 줄, 해시 체인 정상

모든 봉인이 변조 없이 공개 목록과 일치했다. 다만 기준확률이 **11** 건 모두 **0.50** 이라는 점은 한계다. 실제 기출 빈도 같은 근거 있는 기준확률을 쓰면 기술점수가 더 엄격해진다. 다음 봉인부터는 근거 있는 기준확률을 쓰는 것을 과제로 남긴다.

4. 실증 2 — 11 건 채점의 운과 실력(모의)

이 장의 숫자는 모두 모의**(**시뮬레이션**)** 다. 실제 채점 결과가 아니다. 11 건의 확률을 그대로 두고 결과(0/1)만 세 가지 세계에서 20,000 번 무작위로 만들었다.[3]

표 2. 11 건 채점 모의 결과(20,000 회)

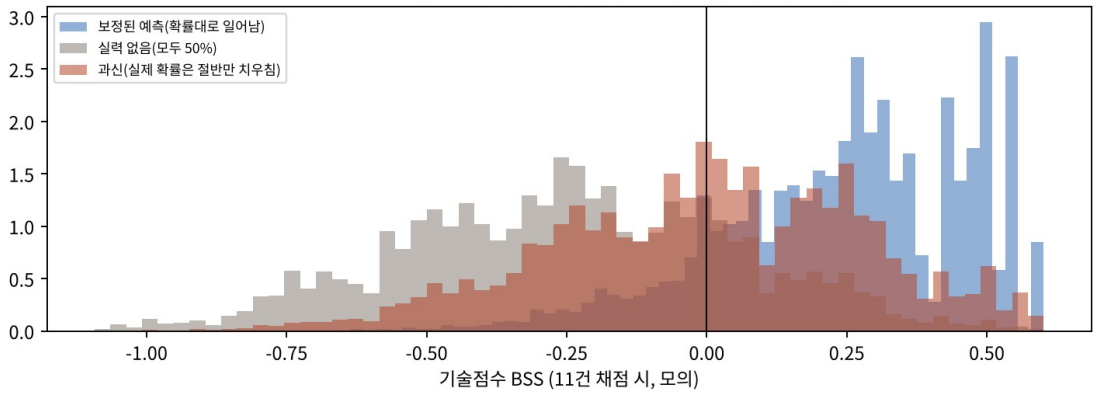
가정한 세계	**BSS** 평균	**BSS>0** 확률	**BPI** 중앙값	배수 중앙값	배수 **95** 분위
보정된 예측(확률대로 일어남)	0.243	85.5%	9.2	1.57	2.21
실력 없음(실제 확률 모두 50%)	−0.247	20.8%	0.0	1.00	1.56
과신(실제	−0.001	51.7%	0.4	1.03	1.99

[2] Brier(1950). 기술점수(skill score) 방식은 기상 예보 검증의 표준이다(Murphy, 1973).

[3] 난수 씨앗 20261005 로 고정했다. 부록 B 의 fd_three.py 2 절로 재현된다.

가정한 세계	**BSS** 평균	**BSS>0** 확률	**BPI** 중앙값	배수 중앙값	배수 **95** 분위
확률은 절반만 치우침)					

그림 4. 11건만 채점했을 때 기술점수가 어디로 갈 수 있나(모의 20,000회)



세 가지가 보인다.

- 보정된 경우에도 결과는 넓게 흩어진다. 확률이 정직하게 맞아도 11 건 중 14.5%의 경우는 기술점수가 0 이하로 나온다. 잘 맞히는 예측자도 첫 성적표가 나쁠 수 있다.
- 실력이 없어도 배수를 받을 수 있다. 실력이 없는 경우의 약 21%가 운으로 배수를 받고, 95 분위에서는 1.56 배까지 오른다. 11 건의 첫 성적만으로 성과 보상을 크게 말하면 안 되는 이유다.
- 과신하면 실력이 **0** 이 된다. 방향은 맞지만 확신이 지나친 예측은 기술점수가 0 근처에 머문다. 확률을 정직하게 매기는 것 자체가 실력이다.

5. 실증 3 — 몇 건이 쌓여야 하나(모의)

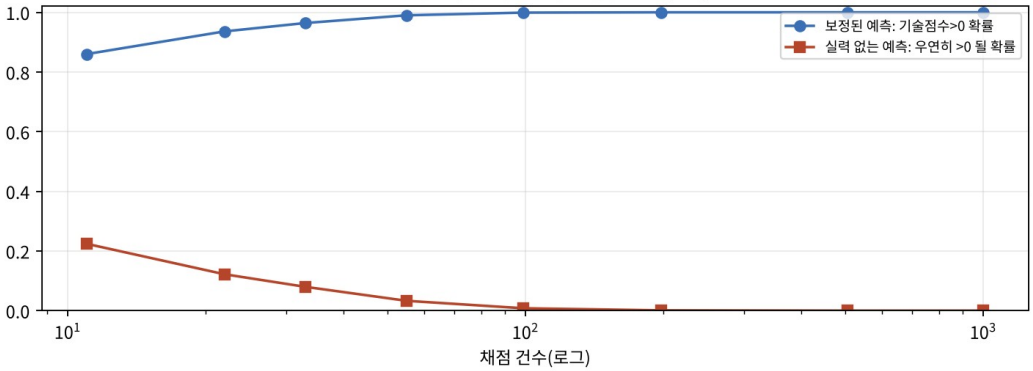
같은 확률 묶음을 반복해 채점 건수를 늘려 가며 6,000 번씩 모의하였다.

표 3. 채점 건수별 운과 실력의 갈림(모의)

채점 건수	보정: **BSS>0**	무실력: **BSS>0**	보정: 배수 중앙값	무실력: 배수 **95** 분위	신뢰계수 **Z**
11	86.0%	22.4%	1.60	1.57	0.355
22	93.6%	12.2%	2.14	1.42	0.524
33	96.4%	8.0%	2.70	1.23	0.623
55	99.0%	3.3%	3.73	1.00	0.733
99	99.9%	0.8%	5.91	1.00	0.832
198	100%	0.1%	10.83	1.00	0.908
506	100%	0.0%	25.93	1.00	0.962

채점 건수	보정: **BSS>0**	무실력: **BSS>0**	보정: 배수 중앙값	무실력: 배수 **95** 분위	신뢰계수 **Z**
1,001	100%	0.0%	50.26	1.00	0.980

그림 5. 채점이 쌓일수록 운과 실력이 갈린다(모의)



실력이 없는 예측자가 우연히 기술점수 양수를 받을 확률은 **55** 건에서 **3%, 99** 건에서 **1%** 아래 로 떨어진다. 이때부터 배수는 사실상 실력만 반영한다. 반대로 보정된 예측자의 배수는 건수에 비례해 커진다. 상한이 없는 실적배수가 **"많이, 오래, 정직하게 맞힌 만큼"** 보상한다는 정책 의도와 맞는 결과다.

5.1 가격 경로(보정된 예측을 가정한 모의 중앙값)

표 4. 모의 가격 경로(가정 — 실제 실적이 아님)

채점 건수	실적배수	**C** 등급**(2,200** 원**)** 제시가	예측 성과 보상
11	1.60	3,500 원	없음
22	2.14	4,700 원	없음
55	3.73	8,200 원	없음
99	5.91	13,000 원	있음(3,100 원)
198	10.83	23,800 원	있음
506	25.93	57,000 원	있음

C 등급 상품이 9,900 원을 넘으려면 이 확률 수준으로 약 99 건이 채점되어야 한다. 봉수의 현재 봉인 속도(논술 한 시즌 11 건)라면 여러 시즌이 걸린다. 성과 보상은 단기간에 생기지 않는다.

6. 논의와 정책 제안

- 소표본 표시. 채점 30 건 미만일 때는 배수 옆에 "소표본 — 운의 영향 큼"을 함께 표시한다. 33 건에서도 실력 없는 예측자의 8%가 양수를 받는다.
- 근거 있는 기준확률. 기준확률 0.50 은 너무 쉬운 기준이다. 대학별 기출 빈도 같은 근거를 기준확률로 봉인하면 기술점수가 실력을 더 정확히 잰다.

- 신뢰계수와 배수의 관계. BPI 에는 신뢰계수 Z 가 들어가지만 실적배수에는 Z 대신 n/5 가 들어간다. 그래서 소표본에서 운의 영향이 배수에 더 크게 남는다. 이를 바로잡는 대안으로는 배수에도 Z 를 곱하거나, 30 건 미만에서는 배수를 반만 적용하는 방법이 있다. 바꿀 때는 새 판으로 공지한다.
- 분야 분리 유지. 논술 실적이 다른 분야 가격을 올리지 않는 지금의 규칙은, 분야마다 표본이 따로 쌓여야 한다는 이 결과와 맞는다.

7. 한계

- 실제 채점 **0** 건: 실제 채점 결과는 0 건이다. 4~5 장은 모두 가정한 세계의 모의다.
- 확률 묶음 반복: 모의는 11 건의 확률 묶음을 반복했다. 실제로는 분야·문항 성격이 달라진다.
- 판정의 오류 가능성: 채점은 AI 가 웹 공개 자료로 하므로 판정 자체가 틀릴 수 있다. 이의가 있으면 확인해 고치고 고친 내역을 공개한다.
- 배수 식의 성격: 실적배수 식(÷5)은 봉수가 정한 제안 기준이며 공인 표준이 아니다.

8. 결론

봉수의 예측 장부는 변조 없이 공개되어 있고, 지금 실적은 0 건이다. 모의 결과는 11 건의 첫 성적이 운에 크게 흔들린다는 것을 보여 준다. 동시에 수십 건이 쌓이면 운과 실력이 분명히 갈리고, 실력이 있으면 배수가 상한 없이 커진다는 것도 보여 준다. 봉수는 첫 채점부터 맞은 것도 틀린 것도 공개하며, 이 논문의 숫자를 실제 성적표로 하나씩 바꿔 나간다.

참고문헌

- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3.
- Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology*, 12(4), 595–600.
- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Tetlock, P. E. and Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*. Crown.
- 봉수 SSGlobal(주) (2026). 공지 — 봉수 표준가격 정책 v1.5 개정(예측 성과 보상). bongsu.kr.

부록 A. 인공지능 이용 고지

본 논문의 계산·모의·그림·초고 문장은 인공지능(Claude)과 함께 작성하였다. 예측 11 건의 문구와 확률은 봉인 원장 그대로이며, 모의 결과는 실제 성적이 아님을 본문 곳곳에 표시하였다. 채점도 AI 가 공개 자료로 하고 근거를 모두 공개한다.

부록 B. 재현 방법

1. 01_로우데이터 에 원장 파일 세 가지를 둔다. 셋 모두 bongsu.kr/jisu 에 공개되어 있다.
   - 봉인예측원장_논술 1_2027_20261005.csv
   - 공개용_봉인해시목록_20261005.txt
   - 채점근거기록.csv
2. 02_파이썬 에서 python -X utf8 fd_three.py 를 실행한다. 결과는 outputs 의 BPI_모의_11건.csv, BPI_모의_표본크기.csv 에 생긴다.
3. 해시 검증만 하려면 bongsu_index.py 의 seal_hash() 로 각 행을 다시 계산해 봉인해시 칸과 비교한다.