

Reddit 소셜미디어 데이터를 활용한 밈주식의 Contrarian Effect 분석

김건희, 최서빈, 황유진
서강대학교

gunhee0928@sogang.ac.kr, implement@sogang.ac.kr, ughichwang@sogang.ac.kr

Analysis of Contrarian Effects in Meme Stocks Using Reddit Social Media Data

Geonhee Kim, Seobin Choi, Yujin Hwang

Sogang University

요약

본 논문은 Reddit r/wallstreetbets 데이터를 활용하여 밈주식에서 나타나는 contrarian effect를 분석하였다. 2021-2023년 기간 동안 GME, AMC, BB 등 주요 밈주식의 Reddit 관심도와 주가 수익률 간의 관계를 종합적으로 분석한 결과, 일관된 음의 상관관계(-0.198~-0.165)를 발견하였다. 이는 Reddit에서 관심이 급증할 때 오히려 수익률이 하락하는 contrarian effect를 의미한다. Event study, 시간적 패턴 분석, 감정 분석, 네트워크 분석을 통해 이 현상의 강건성을 확인하였으며, 53,187개 Reddit 포스트의 텍스트 분석을 통해 행동재무학적 메커니즘을 실증적으로 검증하였다. 특히 확산 표현이 많을수록 수익률 하락으로 이어지는 'behavioral paradox'를 발견하여 기존 소셜미디어-주식 연구와 차별화된 결과를 제시하였으며, 이를 바탕으로 머신러닝 모델의 성능을 유의미하게 향상시켰으며, 밈주식 시장의 비합리성을 실증적으로 설명할 수 있음을 드러내었다.

I. 서론

2021년 GameStop 사태로 대표되는 밈주식 현상은 소셜미디어가 금융시장에 미치는 영향에 대한 새로운 관점을 요구한다. 기존 연구들은 소셜미디어 게시물과 주가상승의 양의 상관관계를 보고해 왔으나 [1~4], 밈주식에서는 소셜미디어에서의 과도한 관심이 오히려 단기 수익률 하락으로 이어지는 contrarian effect가 관찰되어 기존 이론적 틀에 근본적 의문을 제기한다.

본 연구는 Reddit 데이터를 활용하여 이러한 역설적 현상을 체계적으로 규명하고, 이를 행동재무학적 이론 [5~6]을 결합한 프레임으로 메커니즘을 설명해보고자 한다. 나아가 Reddit 기반 신호를 다양한 머신러닝 모델에 통합하여 기존 선형 접근으로 포착되지 않던 비선형 패턴을 검증하고, contrarian effect를 활용한 실증적 예측 모델을 구축함으로써 밈주식 시장의 비합리성을 예측 가능하게 함으로써 이를 밈주식 접근 전략으로서 전환하는 가능성을 제시한다.

II. 본론

2.1 연구방법

데이터 수집: Yahoo Finance API를 통해 2021년 1월부터 2023년 12월까지 GME, AMC, BB의 일간 주가 데이터를 수집하였다. Reddit API를 활용하여 r/wallstreetbets 커뮤니티의 게시물 및 댓글 데이터를 수집하였으며, 총 53,187개의 포스트와 5,409개의 일간 관측치를 분석에 사용하였다.

주요 변수 정의: Reddit Surprise RS는 예상 대비 초과 언급량을 나타내는 핵심 지표로 다음과 같이 정의된다:

$$RS_{i,t} = \frac{ActualMentions_{i,t} - ExpectedMentions_{i,t}}{ExpectedMentions_{i,t}}$$

여기서 i 는 종목, t 는 시점을 나타낸다. 수익률은 일간 수익률로 정의하였다:

$$Returns_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}}$$

여기서 $P_{i,t}$ 는 종목 i 의 t 시점 종가이다.

분석방법: (1) Pearson 상관관계 분석을 통한 Reddit surprise와 수익률 간 관계 분석, (2) Event Study를 통한 극한 Reddit 스파이크 전후 수익률 패턴 분석, (3) 시간적 패턴 분석, (4) 53,187개 포스트에서 행동재무학적 편향 패턴 추출을 위한 텍스트 분석을 수행하였다.

2.2 분석결과

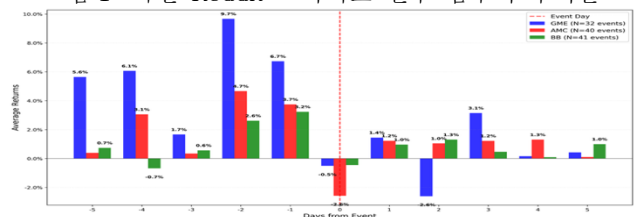
주요 밈주식에서 일관된 음의 상관관계를 발견하였다. GME의 reddit_surprise vs returns_1d = -0.198, AMC = -0.178, BB = -0.165로 모든 종목에서 통계적으로 유의한 음의 상관관계가 나타났다. 이는 Reddit에서 관심이 급증할 때 오히려 다음날 수익률이 하락함을 의미한다.

표 1: Reddit Surprise와 수익률의 상관관계

Stock	Correlation
GME	-0.198
AMC	-0.178
BB	-0.165

Event Study 결과: 극한 Reddit 스파이크(상위 5%) 113개 이벤트를 분석한 결과, 이벤트 당일 평균 음수 수익률(-0.005~-0.026), 다음날 부분적 회복하지만 여전히 저조, 2일 후 추가 조정이 지속되는 패턴을 관찰하였다.

그림 1: 극한 Reddit 스파이크 전후 밈주식 수익률



53,187개 Reddit 포스트의 텍스트 분석을 통해 'Behavioral Paradox'를 발견하였다. 특히 확산 표현이 많을수록 다음날 수익률이 유의하게 하락하는 과잉확신 역설(confidence_score vs next_day_returns:-

0.205, $p=0.0002$)과 언급된 가격에 투자자들이 동조되는 Price Anchoring Effect(price_mentions vs volatility: +0.472)와 FOMO Reversal(urgency_score vs reddit_surprise: -0.127) 효과가 관찰되어 Reddit 관심 급증이 체계적인 단기 가격 하락을 예측하는 contrarian indicator로 작동함을 실증하였다.

2.3 메커니즘 해석

밈주식의 contrarian effect는 소셜미디어 언급 급증이 정보 개선이 아닌 행동적·미시구조적 충격으로 작용하는 체계적 메커니즘을 통해 설명된다. Reddit 언급 폭증은 투자자들의 주의를 특정 종목에 급격히 집중시켜 동조화된 거래행동을 유발하며, 이는 과잉확신 편향과 FOMO 심리가 결합되어 비합리적 거래량 폭증과 단기 과매수를 초래한다. 이 과정에서 호가 스프레드 확대와 시장심도 축소 등 미시구조적 마찰이 발생하고, 높은 공매도 비용과 소음거래자 위험으로 인해 합리적 투자자의 즉각적 차익거래가 제한되어 시장 비효율성이 지속되며, 결국 관심이 분산되고 유동성이 정상화되는 2-3일 후 지연된 가격 조정이 실현되어 Reddit 언급 급증 이후 체계적인 음의 수익률이 나타나는 “예측 가능한 비합리성” 패턴을 형성한다.

2.4 머신러닝 적용

그림 2: Reddit 적용 전후 모델 성능 변화

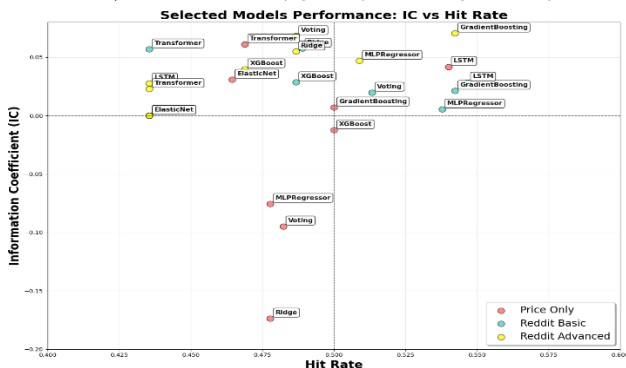


표 2: IC, 적중률 및 ICIR 지표를 통한 모델 성능 비교

Model	IC	Hit Rate	ICIR
Transformer_price	0.061	0.469	0.661
GradientBoosting_price	0.007	0.500	0.513
Voting_adv	0.068	0.487	0.609
Ridge_basic	0.057	0.489	0.468
Transformer_basic	0.057	0.436	-0.453
GradientBoosting_adv	0.070	0.542	0.753

Contrarian effect와 overconfidence paradox의 실증 분석을 넘어 Reddit 기반 신호가 실제 예측력에 기여하는지를 검증하기 위하여 다양한 머신러닝 모델을 적용하였다. 실험 설계는 가격 데이터만을 활용한 Price Only 모델을 baseline으로 설정하고, Reddit 기반 contrarian 신호를 추가한 Reddit Basic 및 Reddit Advanced 모델과 비교하는 방식으로 수행하였다.

Reddit 데이터 기반의 contrarian effect가 적용된 모델들은 전반적으로 price only 모델에 비해 IC와 Hit Rate가 개선된 양상을 보였다. 특히 Gradient Boosting 모델은 Reddit Advanced 피처를 적용했을 때 IC=0.070, Hit Rate=0.542, ICIR=0.753을 기록하며, baseline(GradientBoosting_price_only, IC=0.007, Hit Rate = 0.5) 대비 현저한 향상을 나타냈다.

이러한 결과는 Reddit 기반 contrarian 신호가 단순한 설명 변수에 그치지 않고, 실제 밈 주식 단기 수익률을 예측하는 데 기여할 수 있음을 의미한다. 특히 Gradient Boosting과 같은 비선형 앙상블 모델에서 가장 큰 개선 효과가 관찰된 것은 Reddit 관심 급증이 단순히 투자 심리의 반영을 넘어 예측 가능한 역행적 패턴으로 작동함을 머신러닝 수준에서 실증적으로 확인한 것이다.

III. 결론

본 연구는 Reddit 데이터를 통해 밈주식의 contrarian effect를 발견하고 주의집중 이론과 과잉확신 편향을 중심으로 한 행동재무학 이론으로 이를 설명하였다. 소셜미디어 언급량 급증이 행동적 미시구조적 충격으로 작용하여 단기 주가 하락을 유발함을 실증하였으며 이는 효율적 시장가설의 한계를 보여준다.

핵심 기여점은 네 가지이다. 첫째, 소셜미디어 관심도와 수익률 간 일관된 음의 상관관계라는 기존 연구와 차별화된 현상을 발견하였다. 둘째, 투자자의 확신 수준과 발언 강도 간 역설적 관계를 통해 ‘행동적 역설’이라는 개념을 제시하였다. 셋째, 주의 재배분, 집단조정, 차익거래 제약이라는 세 가지 메커니즘을 통해 현상을 체계적으로 설명하였다. 넷째, Reddit 데이터 적용으로 머신러닝 모델의 성능을 개선시켜, 소셜미디어 데이터가 단순 노이즈가 아닌 실질적 예측 신호임을 입증하였다.

본 연구의 발견은 디지털 플랫폼이 금융시장에 미치는 영향에 대한 이해를 확장하며 투자자 교육과 시장 규제 정책 수립에 시사점을 제공한다. 향후 연구에서는 다양한 소셜미디어 플랫폼과 시장 조건에서의 일반화 가능성을 검증하고 장기적 관점에서의 시장 효율성 회복 과정을 보다 정밀하게 분석할 필요가 있다.

ACKNOWLEDGMENT

본 연구는 서강대학교의 국제지역문화원의 연구비 지원을 받아 수행되었습니다.

참 고 문 헌

- [1] Bollen, J., Mao, H., & Zeng, X. "Twitter mood predicts the stock market," Journal of computational science, vol. 2, no. 1, pp. 1-8, 2011.
- [2] Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. "The effects of Twitter sentiment on stock price returns," PloS one, vol. 10, no. 9, e0138441, 2015.
- [3] Mittal, A., & Goel, A. "Stock prediction using twitter sentiment analysis," Stanford University CS229 Machine Learning Final Project, pp. 1-5, 2012.
- [4] Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. "Tracking retail investor activity," The Journal of Finance, vol. 76, no. 5, pp. 2249-2305, 2021.
- [5] Barber, B. M., & Odean, T. "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors," The review of financial studies, vol. 21, no. 2, pp. 785-818, 2008.
- [6] Daniel, K., Hirshleifer, D., & Subrahmanyam, A. "Investor psychology and security market under-and overreactions," The journal of Finance, vol. 53, no. 6, pp. 1839-1885, 1998.