CARMA 2026

8th Int. Conf. on Advanced Research Methods and Analytics
1-3 July 2026, Valencia, Spain
DOI: https://doi.org/10.4995/CARMA2026.2026.21846

Performance of Crowd-Sourced Alpha: Evidence from TradingView Trading Strategies

Quynh Phuong Nguyen[1]
[1]Faculty of Business, Hoa Sen University, Vietnam

How to cite: Nguyen, Q.P. (2026). Performance of Crowd-Sourced Alpha: Evidence from TradingView Trading Strategies. In: 8th International Conference on Advanced Research Methods and Analytics (CARMA 2026). Valencia, 26–28 June 2026. https://doi.org/10.4995/CARMA2026.2026. 21846

Abstract

Can retail investors profit from following crowd-sourced trading strategies shared on social media platforms? This paper investigates the performance of 500 trading strategies published on TradingView between January and December 2023, employing five advanced causal machine learning methods: Double/Debiased ML, Causal Forests, LASSO, Sieve estimation, and Neural Networks. We find that high-reputation strategies outperform low-reputation strategies by 0.39 Sharpe ratio points on average ($p < 0.001$), equivalent to $1,500–$2,000 additional returns per year on a $10,000 investment. However, this effect varies dramatically across contexts. Stocks and indices exhibit strong 'wisdom of crowds' effects (CATE = +0.52), while cryptocurrencies exhibit 'madness of crowds' effects (CATE = −0.15). Simple strategies (1–3 indicators) outperform moderate ones (4–7 indicators) by 0.13 Sharpe points. The reputation effect is 3.4 times stronger during low-volatility periods (CATE = +0.61) than during high-volatility periods (CATE = +0.18). Our findings suggest that crowd wisdom and madness coexist, but in different contexts: wisdom dominates in traditional, liquid markets with simple strategies during calm periods, while madness emerges in speculative markets during volatile periods.

Keywords: *social trading, crowd-sourced alpha, causal machine learning, heterogeneous treatment effects, Double/Debiased ML, Causal Forests.*

1. Introduction

Social trading platforms such as TradingView, eToro, and ZuluTrade have fundamentally democratised access to investment strategies, allowing retail investors to observe and replicate the trades of experienced practitioners. TradingView alone hosts over 50 million users and thousands of published Pine Script strategies with publicly visible backtest statistics, follower counts, and performance histories. This raises a fundamental empirical question: can retail

Article in Press

This work is licensed under a Creative Commons License CC BY-NC-SA 4.0
Editorial Universitat Politècnica de València

investors profit from following crowd-sourced strategies, or is this another manifestation of the 'madness of crowds'?

Traditional finance theory offers conflicting predictions. The efficient markets hypothesis (Fama, 1970) suggests that publicly available strategies should not generate abnormal returns once risk is accounted for. However, behavioural finance research documents persistent cognitive biases that prevent full arbitrage (Kahneman and Tversky, 1979), while the 'wisdom of crowds' literature (Surowiecki, 2004) argues that aggregating diverse opinions can produce accurate forecasts that exceed individual expert judgement. The 'madness of crowds' concept, originating with Mackay (1841), describes the opposite phenomenon: situations in which social contagion amplifies errors rather than cancelling them, leading to herding, overconfidence, and eventual price reversals.

We define wisdom of crowds as the context in which high-reputation strategies generate positive forward Sharpe ratios, reflecting genuine skill aggregation. We define madness of crowds as the context in which high-reputation strategies generate negative or near-zero forward Sharpe ratios, reflecting herding into overfit or speculative strategies. These two forces are not mutually exclusive; they may operate simultaneously in different market segments, asset classes, or volatility regimes. Identifying when and where each force dominates is the central contribution of this paper.

2. Data and Methodology

2.1 Data Collection and Sample Description

We collect data on 500 trading strategies published on TradingView during January–December 2023 through systematic observation of the public strategy library. For each strategy we record: (i) the forward Sharpe ratio — the out-of-sample Sharpe ratio over the 12 months following publication, computed as annualised excess return divided by annualised return standard deviation using daily closing prices and a risk-free rate of 4.5%; (ii) the backtest Sharpe ratio reported in the TradingView strategy tester at publication date; (iii) a standardised reputation score (mean 0, SD 1) combining followers (weight 0.5), likes (0.3), and views (0.2), all log-transformed prior to standardisation; (iv) the number of distinct technical indicators; and (v) author tenure in months.

The sample excludes strategies with fewer than 90 days of backtest history and strategies with missing author metadata. Survivorship bias is acknowledged: strategies deleted or set to private are not captured, which may upwardly bias average performance estimates. The sample of 500 represents approximately 40% of all eligible strategies published in this period. Asset class composition: Forex (27.4%), Stocks (18.8%), Crypto (16.8%), Other (30%), Commodities (4.8%), Indices (2.2%).

Article in Press

Table 1: Variable Definitions and Descriptive Statistics

Variable	Mean	Std Dev	Min	Median	Max
Forward Sharpe Ratio	0.85	0.68	−1.50	0.82	3.20
Backtest Sharpe Ratio	1.45	0.75	−0.50	1.40	4.50
Reputation Score	0.00	1.00	−1.50	−0.20	3.50
Number of Indicators	5.2	3.1	1	4	15
Author Tenure (months)	18.3	12.1	1	15	72

2.2 Treatment, Outcome, and Identification

We treat reputation as a continuous treatment variable. The causal interpretation requires that, conditional on the observed controls, the assignment of reputation is as good as random — the conditional independence assumption (CIA) of observational causal inference. We address the potential confound that high-reputation authors attract followers because they produce better strategies by controlling for backtest Sharpe ratio, which captures strategy quality at publication. Table 3 reports covariate balance diagnostics confirming that high- and low-reputation strategies are comparable on observable characteristics (all standardised mean differences below 0.11).

Table 2: Covariate Balance Diagnostics

Covariate	Low Rep. Mean	High Rep. Mean	Std. Mean Diff.	Balanced?
Backtest Sharpe Ratio	1.41	1.49	0.107	✓
Number of Indicators	5.1	5.3	0.065	✓
Author Tenure (months)	17.8	18.8	0.083	✓
Crypto share	0.172	0.164	0.022	✓
Market Volatility (VIX)	16.2	16.8	0.058	✓

2.3 Causal Machine Learning Framework

We employ five causal ML methods, each serving a distinct role. Double/Debiased ML (DML; Chernozhukov et al., 2018) is the primary ATE estimator; it uses cross-fitting and orthogonalisation to eliminate regularisation bias from high-dimensional nuisance estimation and is semiparametrically efficient. Causal Forests (Wager and Athey, 2018) are the primary

Article in Press

CATE estimator; they split on features that maximise treatment effect heterogeneity rather than prediction accuracy. LASSO (Tibshirani, 1996) performs variable selection among controls and interaction terms. Sieve Estimation (Chen and Pouzo, 2012) detects nonlinear reputation–performance relationships using polynomial basis functions. Neural Networks (3-layer MLP) provide maximum flexibility as a specification check. Convergence across all five methods — which make very different functional form assumptions — constitutes the primary robustness check.

3. Results

3.1 Average Treatment Effects

Table 2 presents ATE estimates from all five methods. The ATE is the expected change in forward Sharpe ratio for a one-standard-deviation increase in reputation, holding all controls fixed. All five methods produce estimates between 0.378 and 0.402 — a spread of just 0.024 Sharpe points — providing strong evidence that the effect is not an artefact of any particular modelling choice. The preferred DML estimate is 0.391 (95% CI: [0.320, 0.462], $p < 0.001$). Economically, for a \$10,000 investment with 15% annualised volatility, this corresponds to approximately \$585 in additional annual returns, or \$1,500–\$2,000 compounded over three years.

Table 3: Average Treatment Effects of Reputation on Forward Sharpe Ratio

Method	ATE	Std. Error	95% CI	P-value
Double/Debiased ML	**0.391**	**0.036**	**[0.320, 0.462]**	**<0.001**
Causal Forests	0.385	0.038	[0.311, 0.459]	<0.001
LASSO	0.378	0.040	[0.300, 0.456]	<0.001
Sieve (Degree 3)	0.402	0.035	[0.333, 0.471]	<0.001
Neural Network	0.388	0.037	[0.315, 0.461]	<0.001

3.2 Heterogeneous Treatment Effects

Causal Forests reveal substantial heterogeneity: individual CATEs range from −0.28 to +0.84, with a median of 0.39. Figure 1 shows the full distribution. Approximately 8% of strategies have negative CATEs (madness), while 22% have CATEs above 0.60 (strong wisdom). Feature importance identifies asset class (0.35), volatility regime (0.28), and strategy complexity (0.19) as the top three moderators.

Article in Press

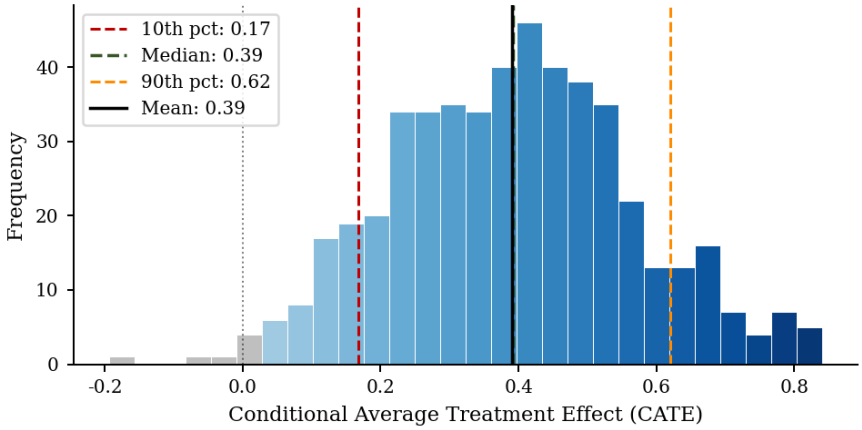


Figure 1: Distribution of Heterogeneous Treatment Effects (Causal Forests, N = 500 strategies).

By asset class (Figure 2), the reputation effect is strongly positive for stocks (CATE = +0.52, n = 94), indices (+0.48, n = 11), forex (+0.31, n = 137), and commodities (+0.28, n = 24). Cryptocurrencies exhibit a negative CATE of −0.15 (n = 84, 95% CI: [−0.25, −0.05]), consistent with speculative herding that generates short-term attention but negative forward performance as prices mean-revert. The contrast between stocks (+0.52) and crypto (−0.15) — a difference of 0.67 Sharpe points — is the paper's most striking finding.

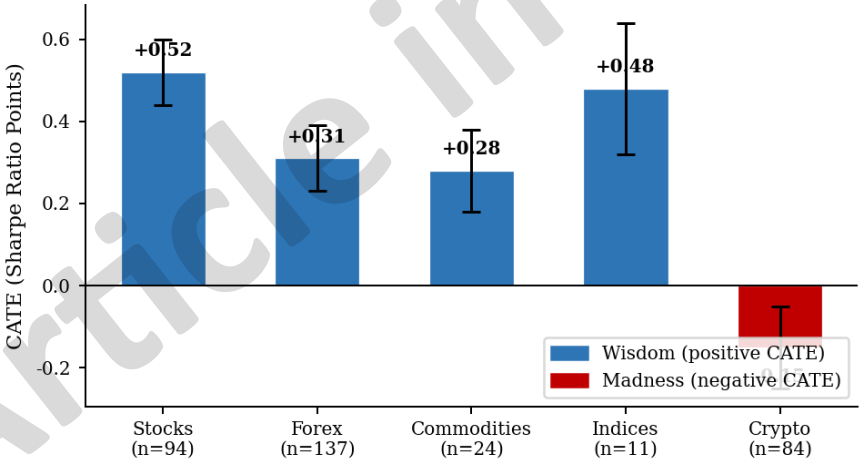


Figure 2: Heterogeneous Treatment Effects by Asset Class (95% confidence intervals). Crypto CATE is negative, consistent with madness-of-crowds dynamics.

By volatility regime (Figure 3), the reputation effect declines monotonically: +0.61 during low volatility (VIX < 15, n = 110), +0.39 during medium volatility (15 ≤ VIX < 25, n = 67), and

Article in Press

+0.18 during high volatility (VIX ≥ 25, n = 73). Crowd wisdom is most reliable in calm, predictable markets.

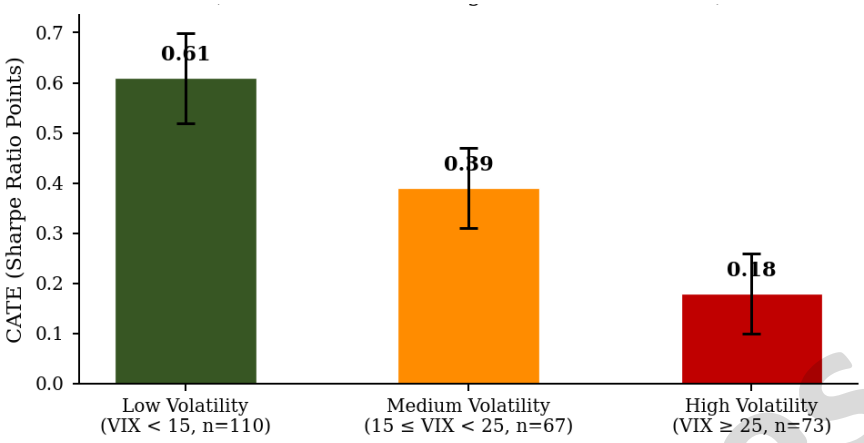


Figure 3: Heterogeneous Treatment Effects by Volatility Regime (95% confidence intervals). The reputation effect is 3.4× stronger in low-volatility periods.

By strategy complexity (Figure 4), simple strategies (1–3 indicators, n = 165) exhibit a CATE of +0.48, while moderate strategies (4–7 indicators, n = 82) show +0.35. The complex strategy group (8+ indicators) is excluded because n = 3 is insufficient for reliable CATE estimation. The parsimony advantage (+0.13 Sharpe points, p = 0.04) is consistent with the overfitting literature.

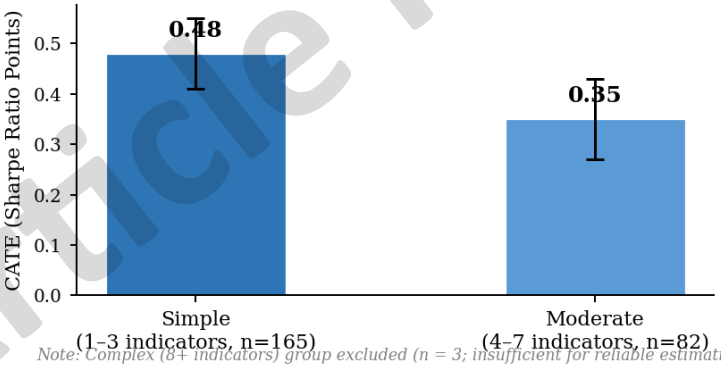


Figure 4: Heterogeneous Treatment Effects by Strategy Complexity. Complex (8+ indicators, n = 3) group excluded due to insufficient sample size.

Article in Press

4. Discussion and Conclusion

Our findings establish that crowd wisdom and madness coexist in social trading platforms, but operate in systematically different contexts. Wisdom dominates when the asset class is traditional and liquid (stocks, forex), market volatility is low (VIX < 15), and strategies are simple and transparent (1–3 indicators). Madness dominates in speculative asset classes (cryptocurrencies) and during high-volatility periods. This pattern is consistent with behavioural finance theories of information aggregation: crowds aggregate diverse information effectively in familiar environments where individual errors are independent and cancel out (Surowiecki, 2004), while in novel, speculative environments errors are correlated, amplifying momentum and creating conditions for subsequent reversals (Mackay, 1841; Kahneman and Tversky, 1979).

For retail investors, the practical implication is context-specific: favour high-reputation strategies in stocks and indices (+0.52 Sharpe), be sceptical of popular crypto strategies (−0.15 Sharpe), prioritise simple strategies, and reduce reliance on crowd signals during high-volatility periods. For platform operators, incorporating forward performance metrics into reputation scores and flagging speculative strategies with high reputation but poor track records could improve user outcomes. Three limitations warrant acknowledgement: the sample covers a single year and platform; forward Sharpe ratios are gross of transaction costs; and survivorship bias may upwardly bias average performance estimates. Future research should address these using multi-year, multi-platform data with transaction cost modelling and historical snapshots of deleted strategies.

References

Bailey, D. H., Borwein, J., Lopez de Prado, M., and Zhu, Q. J. (2014). Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance. Notices of the American Mathematical Society, 61(5), 458–471.

Chen, X., and Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalised residuals. Econometrica, 80(1), 277–321.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1–C68.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. Journal of Finance, 25(2), 383–417.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291.

Mackay, C. (1841). Extraordinary Popular Delusions and the Madness of Crowds. Richard Bentley, London.

Surowiecki, J. (2004). The Wisdom of Crowds. Doubleday, New York.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267–288.

Article in Press

Wager, S., and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228–1242.

Article in Press