ALGORITHMIC TRADING, QUANTITATIVE FINANCE

Do Simple Trading Rules Survive Out of Sample? Evidence on Backtest Overfitting and Trading Costs from Indian Equity Indices, with Lessons for No-Code Platforms

Ayush Yadav, Bhojraj Shewale

Amity Business School, Amity University Mumbai, Mumbai, India

Published November 2026

Abstract

No-code and AI-based backtesting platforms let retail investors in India test trading ideas on past data in minutes. This makes testing cheap, but it also makes it easy to try many variations and keep the one that looks best. This paper asks how far simple technical rules hold up when they are checked on data they were not chosen on, and how much trading costs change the picture. I test 74 long-or-cash rules from four common families (moving average crossovers, price above a moving average, breakouts and RSI mean reversion) on daily Nifty 50 and Nifty Bank index data from February 2012 to November 2026. Rules are ranked on 2012 to 2019 data and checked on 2020 to 2026 data, with a cost of 0.10% per side. On the Nifty 50, only one of the 74 rules beat buy and hold on Sharpe ratio in-sample. The in-sample and out-of-sample rankings were fairly closely linked (rank correlation 0.69), helped by one large event, the March 2020 crash. On Nifty Bank, the best in-sample rule saw its Sharpe ratio fall from 1.08 to 0.37, below buy and hold, and only 7 of 74 rules beat buy and hold out of sample. A deflated Sharpe ratio of 0.77 to 0.82 and an estimated probability of backtest overfitting of about 0.35 show that the best in-sample results are not reliable. One thousand random rules produced an in-sample Sharpe ratio of 1.17, higher than any real rule, which fell to 0.29 out of sample. A walk-forward version that re-selects the best moving average rule each year earned 5.7% a year against 12.3% for buy and hold. Costs had a large effect on rules that trade often. The results suggest that backtesting platforms aimed at retail users should report out-of-sample results, trading costs, a buy-and-hold benchmark and the number of variations tried as standard features.

Keywords: backtesting; overfitting; technical analysis; Nifty 50; retail investors; algorithmic trading; deflated Sharpe ratio; India

JEL classification: G11; G14; G17

1 Introduction

Retail participation in Indian equity markets has grown very quickly. The National Stock Exchange (NSE) crossed 13 crore unique registered investors in April 2026, and its count of trading accounts crossed 26 crore in June 2026 (NSE, 2026). The median investor is now about 33 years old. At the same time, the Securities and Exchange Board of India (SEBI) has found in study after study that most individual traders in equity derivatives lose money. In FY26, 87.7% of individual traders made net losses, with total losses of ₹91,685 crore (SEBI, 2026).

Over the same period, trading has become more automated. Algorithms now place most trades on NSE, and retail traders have gained access to broker APIs and third-party platforms. SEBI's framework for retail algorithmic trading, issued in February 2025, applies to all brokers from 1 April 2026 (SEBI, 2025). A new group of tools has also appeared: no-code and AI-based backtesting platforms. These let a user describe a strategy using menus or plain English and see how it would have performed on past data.

Backtesting is a sensible first step before risking money. But it has a well-known weakness. If a user tries many variations of a rule on the same data, some will look good by chance alone. Bailey et al. (2014) showed that with enough trials, a strategy with an impressive Sharpe ratio is almost certain to appear even when no variation has any real edge. Tools that make testing fast and free make this problem worse, because trials cost nothing.

This paper studies the problem using Indian index data and the kind of simple rules a retail user is likely to test first. It asks four questions.

1. 1. Do simple technical rules beat buy and hold on the Nifty 50 and Nifty Bank after a realistic trading cost?
2. 2. Does the rule that looks best on past data stay among the best on later data?
3. 3. How much do trading costs change the results?
4. 4. What should a retail backtesting report show so that it does not mislead its users?

The paper makes three contributions. First, it gives recent, out-of-sample evidence on simple rules for two major Indian indices, covering the period up to November 2026. Second, it applies formal tools for measuring overfitting, the deflated Sharpe ratio and the probability of backtest overfitting, to a retail-style rule set. Third, it links these results to the design of retail backtesting platforms, a topic that has received little attention in the research literature.

The rest of the paper is organised as follows. Section 2 reviews the literature. Section 3 describes the data and method. Section 4 presents the results. Section 5 discusses what they mean for investors and for backtesting platforms. Section 6 sets out the limitations, and Section 7 concludes.

2 Literature review

2.1 Technical trading rules

The efficient market hypothesis holds that prices reflect available information, so past prices should not help predict future returns (Fama, 1970). Under this view, technical rules should not beat buy and hold after costs. Brock et al. (1992) challenged this view with evidence that moving average and trading range rules seemed to predict returns on the Dow Jones index over almost a century. Their results drew wide attention and led to a large body of follow-up work.

Park and Irwin (2007) reviewed 95 modern studies and found that more than half reported positive results for technical analysis. They also noted that many studies did not deal properly with data snooping, risk and costs. Evidence for emerging markets has been mixed. Gunasekarage and Power (2001) found that moving average rules had some predictive power in four South Asian markets, including India. Mitra (2011) studied the Indian market and found that the gains from technical rules became small once realistic costs were included.

Trend-following rules have stronger support across asset classes. Faber (2007) showed that holding an asset only when it is above its 10-month moving average reduced drawdowns without a large loss of return in long US samples. Moskowitz et al. (2012) documented time series momentum in a wide range of futures markets. Lo (2004) proposed the adaptive markets hypothesis, in which the profit from a simple rule can appear and fade as market conditions and participants change.

2.2 Data snooping and backtest overfitting

Sullivan et al. (1999) revisited the rules of Brock et al. (1992) using the reality check test of White (2000), which adjusts for the fact that the best rule was chosen from a large set. They found that the best rule did well in the original sample but did not keep working on later data. This was an early and clear example of how data snooping can create false discoveries.

More recent work focuses on the backtest itself. Bailey et al. (2014) showed how easily backtests can be overfitted and argued that results are meaningless unless the number of trials is reported. Bailey and López de Prado (2014) proposed the deflated Sharpe ratio, which estimates the probability that a strategy's true Sharpe ratio is above zero after allowing for the number of trials, the sample length and the skewness and kurtosis of returns. Bailey et al. (2017) proposed combinatorially symmetric cross-validation (CSCV) to estimate the probability of backtest overfitting (PBO): the chance that the best in-sample strategy performs below the median out of sample. Harvey and Liu (2015) argued that a Sharpe ratio must clear a much higher bar when many strategies have been tried, and Harvey et al. (2016) made the same point for academic research on return factors. López de Prado (2018) summed up this view by arguing that a backtest should confirm a sound idea, not search for one.

2.3 Retail investors and trading tools

Barber and Odean (2000) showed that US households that traded most earned the lowest net returns. Barber et al. (2014) found that fewer than one percent of day traders in Taiwan earned predictable profits after fees. SEBI's studies of Indian derivatives traders point in the same direction. Its FY26 study found that before transaction costs 82.1% of individual traders had losses, rising to 87.7% after costs (SEBI, 2026).

Research on fintech adoption, often based on the technology acceptance model of Davis (1989), finds that ease of use and usefulness drive the take-up of new tools. No-code backtesting platforms apply this directly by removing the need to code. However, I found no published study that connects the overfitting literature to the design of retail backtesting tools in India. This paper aims to start filling that gap.

3 Data and methodology

3.1 Data

The study uses daily closing values of the Nifty 50 and Nifty Bank indices from 21 February 2012 to 13 November 2026, a total of 3,640 trading days for each index. The data was taken from the open eod2 dataset (Thadikaran, 2026), which compiles end-of-day data from NSE's published reports. The index values were checked against known year-end levels. For example, the Nifty 50 closed 2019 at 12,168 and 2020 at 13,982 in the data, matching published values, and the largest one-day fall, of 13.0% on 23 March 2020, matches the known COVID crash.

Both series are price indices. They do not include dividends, which lowers the measured return of buy and hold by roughly one to one and a half percentage points a year compared with a total return index. Time spent out of the market earns no interest. These two choices work in opposite directions and are discussed in Section 6.

3.2 Sample split

The sample is split into two periods.

- In-sample: 21 February 2012 to 31 December 2019 (1,932 trading days), used to rank and select rules.
- Out-of-sample: 1 January 2020 to 13 November 2026 (1,707 trading days), used to check them.

The split was fixed before looking at any results. The out-of-sample period includes the March 2020 crash, the strong recovery of 2020 and 2021, and the more uneven markets of 2024 to 2026.

3.3 Trading rules

I test 74 rules from four families that are common on retail platforms and in trading education. All rules are long or in cash: they either hold the index or stay out. None use short selling or leverage.

Table 1. Rule families and parameters

Family	Rule	Parameters tested	Number of rules
Moving average crossover (SMA)	Hold when the fast moving average is above the slow moving average	Fast: 5, 10, 15, 20, 30, 40, 50 days; slow: 50, 75, 100, 150, 200 days (fast below slow)	34
Trend (price above SMA)	Hold when the closing price is above its moving average	20, 50, 100, 150, 200, 250 days	6
Breakout	Enter when the close is above the highest close of the last n days; exit when below the lowest close of the last m days	n: 20, 55, 100; m: 10, 20, 50	9
RSI mean reversion	Enter when 14-day RSI falls below a lower level; exit when it rises above an upper level	Lower: 20, 25, 30, 35, 40; upper: 50, 55, 60, 65, 70	25

Source: Author

3.4 Backtest mechanics

A signal calculated at the close of day t sets the position for day t+1. This avoids look-ahead bias, since the rule never uses a price it could not have known. The daily return of a rule is the position held multiplied by the index return, minus a cost on every change of position:

$$r(\text{strategy}, t) = \text{position}(t-1) \times r(\text{index}, t) - c \times |\text{position}(t-1) - \text{position}(t-2)|$$

where c is the cost per side. The base case uses c = 0.10%, which allows for brokerage, securities transaction tax, exchange charges and slippage when trading an index fund or ETF. Costs of 0%, 0.05%, 0.20% and 0.30% are also tested.

3.5 Performance measures

For each rule and period I calculate the compound annual growth rate (CAGR), annualised volatility, the annualised Sharpe ratio (mean daily return divided by its standard deviation, multiplied by the square root of 252, with a zero risk-free rate), maximum drawdown and the number of round-trip trades. Buy and hold is the benchmark.

3.6 Measures of overfitting

Four checks are used.

a. Rank correlation: the Spearman correlation between in-sample and out-of-sample Sharpe ratios across all 74 rules. A high value means the in-sample ranking carries information about the future.

b. Deflated Sharpe ratio (DSR): following Bailey and López de Prado (2014), the expected maximum Sharpe ratio from N independent trials is estimated as

$$E[\max SR] = \sigma(SR) \times [(1 - \gamma) \times Z(1 - 1/N) + \gamma \times Z(1 - 1/(N \times e))]$$

where $\sigma(SR)$ is the standard deviation of Sharpe ratios across trials, γ is the Euler-Mascheroni constant (about 0.5772) and Z is the inverse of the standard normal distribution. The DSR is the probability that the best rule's true Sharpe ratio exceeds this level, allowing for the skewness and kurtosis of its returns. A DSR above 0.95 is usually taken as reliable.

c. Probability of backtest overfitting (PBO): following Bailey et al. (2017), the full sample is divided into 16 equal blocks. For 2,000 randomly chosen ways of splitting these blocks into two halves, the best rule on the first half is found and its rank on the second half is recorded. The PBO is the share of splits in which the selected rule ranks at or below the median.

d. Random rules: 1,000 rules that switch between the index and cash at random, with an average holding of about 30 days in the market and 16 days out, giving an exposure similar to the trend rules. Their results show how good a rule can look by luck.

Finally, a walk-forward test re-selects, at the start of each year from 2017 to 2026, the moving average crossover with the best Sharpe ratio over the previous five years, and holds it for that year. This mimics a user who keeps re-optimising.

4 Results

4.1 Buy and hold and selected rules

Table 2 shows buy and hold and six well-known rules over the full sample at the base cost.

Table 2. Nifty 50, full sample (February 2012 to November 2026), cost 0.10% per side

Rule	CAGR	Volatility	Sharpe ratio	Maximum drawdown	Round trips
Buy and hold	10.9%	16.0%	0.73	-38.4%	1
SMA 10/75	9.4%	10.7%	0.89	-19.9%	30
SMA 50/200	3.7%	11.8%	0.36	-42.7%	10
Price above SMA 200	4.4%	10.9%	0.45	-26.0%	63
Price above SMA 20	6.2%	10.0%	0.65	-17.4%	189

Rule	CAGR	Volatility	Sharpe ratio	Maximum drawdown	Round trips
RSI 30/55	1.1%	9.8%	0.16	-32.7%	21
Breakout 20/10	3.0%	8.4%	0.39	-22.1%	90

Source: Author's calculations

Buy and hold earned the highest CAGR, 10.9% a year. Every timing rule earned less. Most of them also had a lower Sharpe ratio. The exception, SMA 10/75, had a Sharpe ratio of 0.89 and roughly half the drawdown of buy and hold, but as shown below it was the best of 74 rules in-sample, so its result must be judged with care. The widely followed 50/200 "golden cross" did poorly, earning 3.7% a year with a drawdown larger than buy and hold. It tended to exit after much of a fall had happened and to re-enter after much of a recovery.

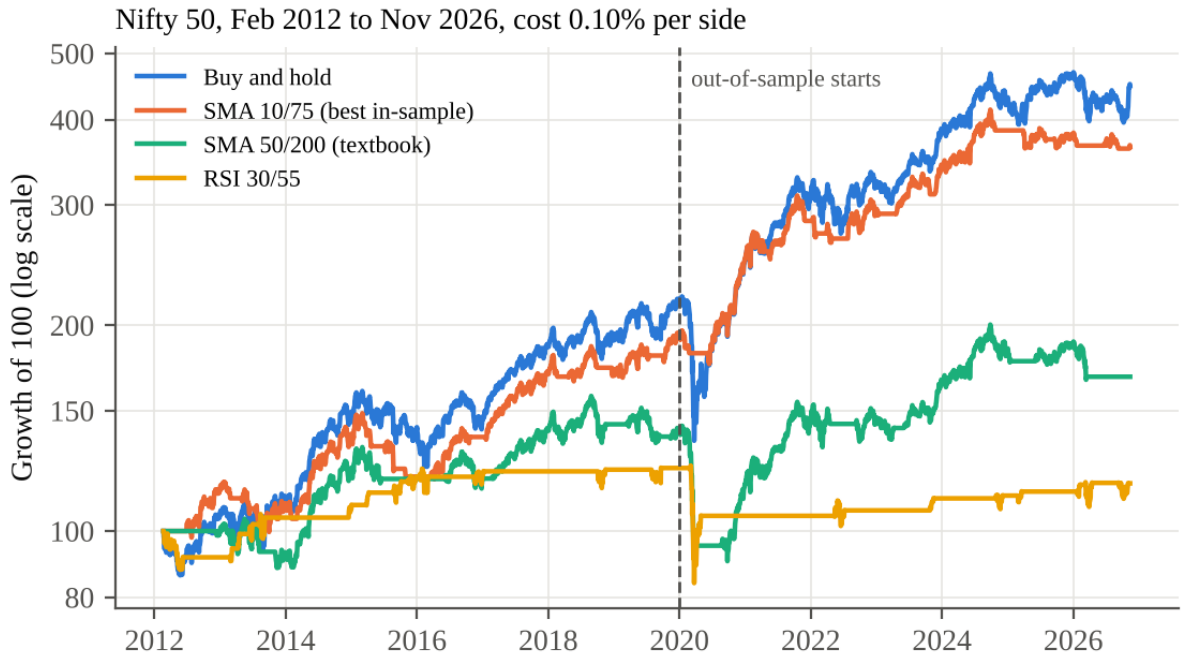


Figure 1. Growth of 100 under buy and hold and three rules, Nifty 50 (log scale)

Source: Author's calculations. Dashed line marks the start of the out-of-sample period.

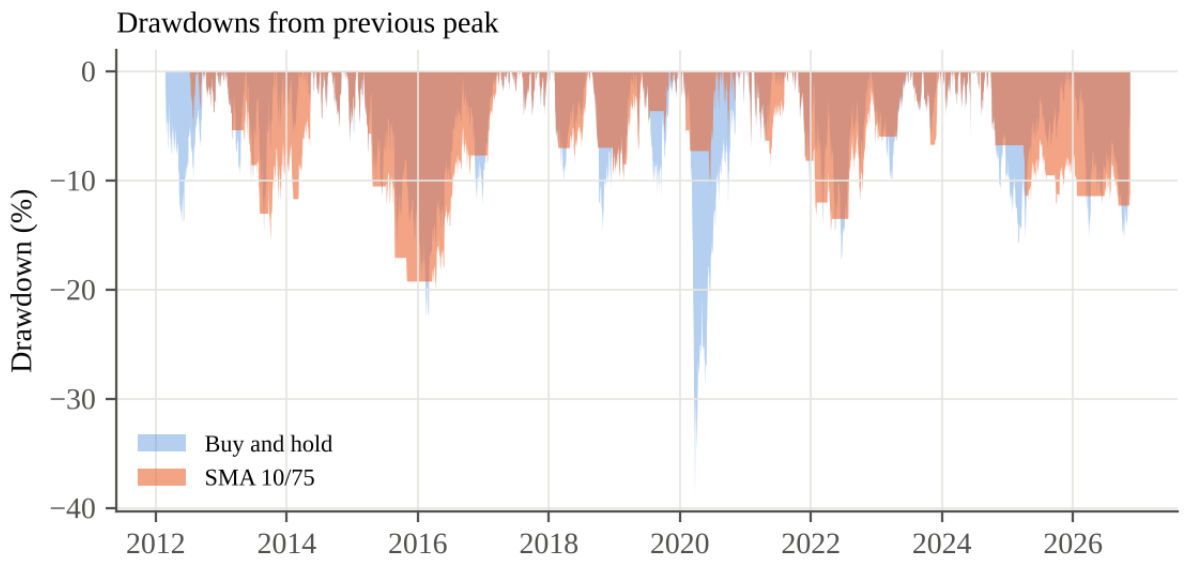


Figure 2. Drawdowns of buy and hold and SMA 10/75, Nifty 50

Source: Author's calculations

Figure 2 shows where the drawdown benefit came from. Most of it came from one episode, the March 2020 crash, when the SMA 10/75 rule was out of the market for much of the fall. In calmer years the rule often lagged, as in 2013, 2015, 2021 and 2025 (Table 3).

Table 3. Calendar-year returns, Nifty 50

Year	Buy and hold	SMA 10/75	Year	Buy and hold	SMA 10/75
2012 (part)	5.2%	14.6%	2020	14.9%	29.4%
2013	6.8%	-4.4%	2021	24.1%	13.7%
2014	31.4%	25.0%	2022	4.3%	3.4%
2015	-4.1%	-12.3%	2023	20.0%	16.6%
2016	3.0%	14.3%	2024	8.8%	12.5%
2017	28.6%	24.7%	2025	10.5%	-1.5%
2018	3.2%	0.4%	2026 (part)	-3.8%	-4.0%
2019	12.0%	12.5%			

Source: Author's calculations, cost 0.10% per side

4.2 In-sample selection and out-of-sample performance

Table 4 summarises the in-sample and out-of-sample results of all 74 rules for both indices.

Table 4. Summary of the 74 rules, in-sample (2012-19) and out-of-sample (2020-26)

Measure	Nifty 50	Nifty Bank
Buy and hold Sharpe ratio, in-sample	0.78	0.76
Buy and hold Sharpe ratio, out-of-sample	0.69	0.52
Rules beating buy and hold Sharpe, in-sample	1 of 74	12 of 74

Measure	Nifty 50	Nifty Bank
Rules beating buy and hold Sharpe, out-of-sample	29 of 74	7 of 74
Median rule Sharpe ratio, in-sample	0.47	0.47
Median rule Sharpe ratio, out-of-sample	0.58	0.29
Best in-sample rule	SMA 10/75	SMA 20/50
Its Sharpe ratio, in-sample	0.87	1.08
Its Sharpe ratio, out-of-sample	0.91	0.37
Its out-of-sample rank (of 74)	7	27
Spearman rank correlation, in-sample vs out-of-sample Sharpe	0.69	0.50

Source: Author's calculations, cost 0.10% per side

The two indices tell different stories. On the Nifty 50, the in-sample period was hard for timing rules: only one of 74 rules beat buy and hold. The out-of-sample period was easier, with 29 rules ahead of buy and hold, and the best in-sample rule kept its Sharpe ratio and ranked 7th. The rank correlation of 0.69 suggests that the in-sample ranking carried real information. However, much of the out-of-sample benefit came from avoiding the March 2020 crash, a single event that may not repeat in the same way.

On Nifty Bank, the in-sample period flattered timing rules: 12 beat buy and hold, and the best rule, SMA 20/50, had a Sharpe ratio of 1.08 against 0.76 for buy and hold. Out of sample, that rule's Sharpe ratio fell to 0.37, below buy and hold at 0.52. Its CAGR fell from 15.3% to 4.6%. Only 7 of 74 rules beat buy and hold. A user who chose a rule from a backtest ending in 2019 would have done worse than simply holding the index.

Table 5. Top five in-sample rules and their out-of-sample results

Index	Rule	In-sample Sharpe	Out-of-sample Sharpe	In-sample CAGR	Out-of-sample CAGR	Out-of-sample rank
Nifty 50	SMA 10/75	0.87	0.91	9.0%	9.8%	7
Nifty 50	Breakout 20/50	0.77	0.76	8.5%	9.5%	22
Nifty 50	SMA 15/75	0.76	1.11	7.6%	12.2%	1
Nifty 50	SMA 30/75	0.72	0.80	7.1%	8.5%	16
Nifty 50	SMA 10/50	0.71	0.91	6.9%	9.8%	6
Nifty Bank	SMA 20/50	1.08	0.37	15.3%	4.6%	27
Nifty Bank	Breakout 20/20	0.96	0.46	13.7%	5.6%	16
Nifty Bank	Breakout 20/50	0.94	0.54	15.0%	8.5%	6
Nifty Bank	Price above SMA 100	0.93	0.28	12.8%	3.1%	38
Nifty Bank	Price above SMA 50	0.92	0.59	13.0%	7.6%	4

Source: Author's calculations, cost 0.10% per side

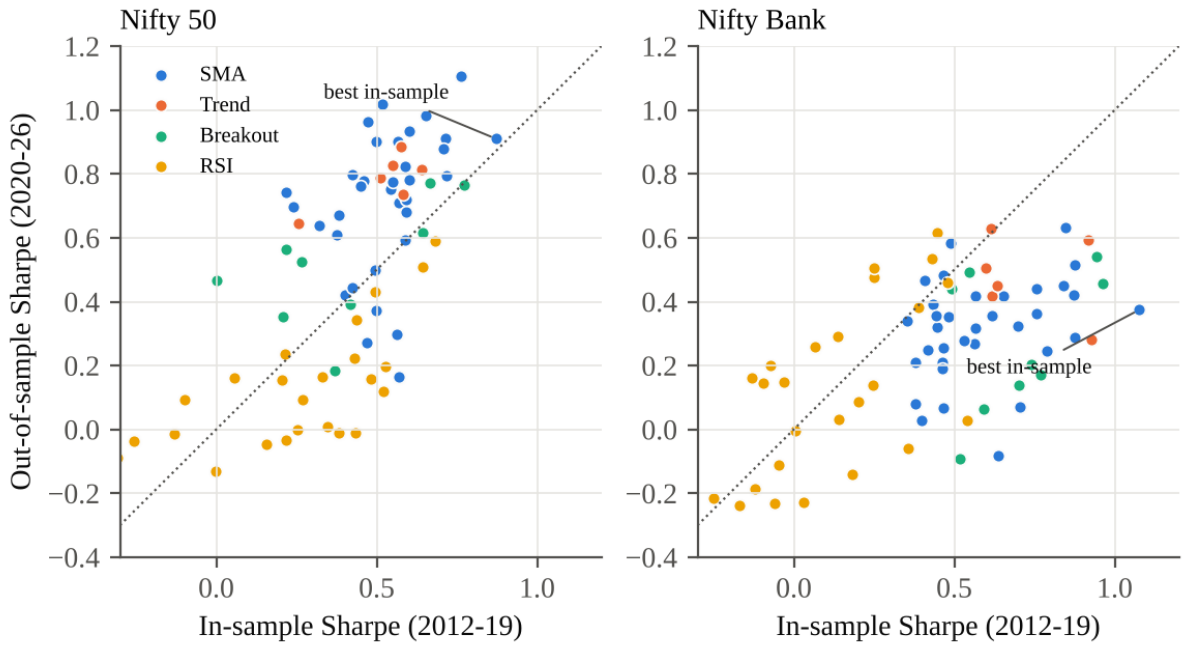


Figure 3. In-sample versus out-of-sample Sharpe ratios of 74 rules

Source: Author's calculations. The dotted line marks equal Sharpe ratios in both periods.

Figure 3 shows the pattern by rule family. On the Nifty 50, most moving average and trend rules lie above the dotted line, meaning they did better out of sample than in-sample, while most RSI rules lie near or below zero. On Nifty Bank, almost all rules lie below the line, and the best in-sample rule sits far below it.

By family, RSI mean-reversion rules were the weakest on both indices, with average Sharpe ratios of 0.24 and 0.12 on the Nifty 50 in the two periods. They were in the market only about 22% of the time and so missed most of the index's rise. Trend and crossover rules, which were in the market about two thirds of the time, did best.

4.3 Sensitivity to parameters

Figure 4 shows the Sharpe ratios of all 34 moving average crossovers on the Nifty 50 as a grid. A sound rule should work across nearby parameters, not just at one point.

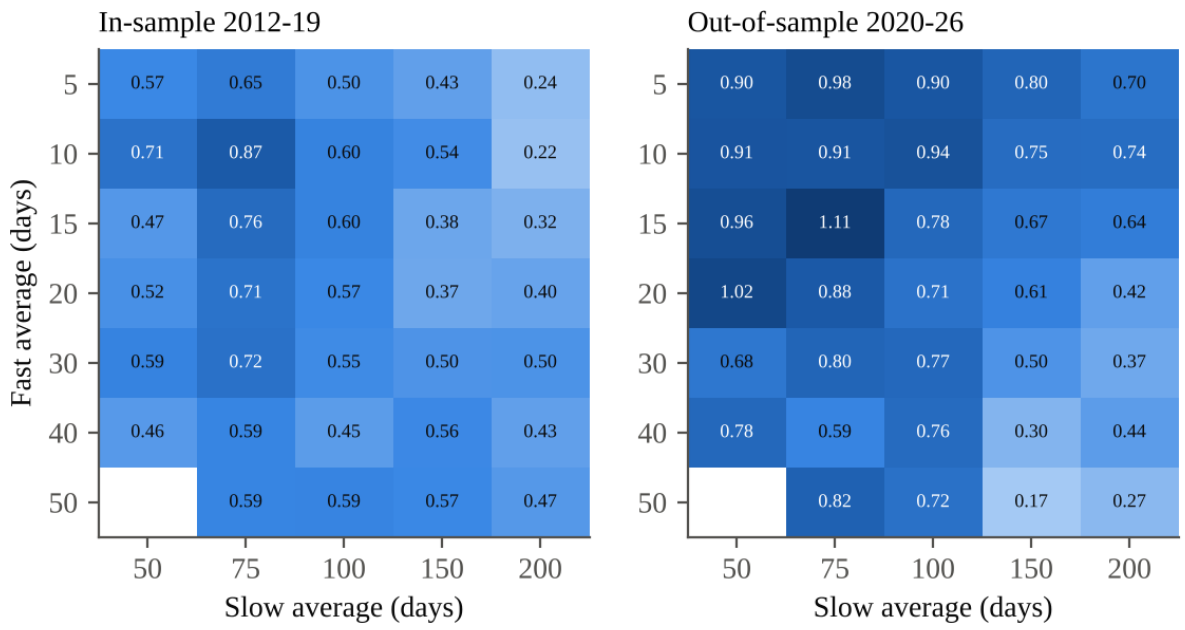


Figure 4. Sharpe ratios of moving average crossovers by fast and slow length, Nifty 50

Source: Author's calculations, cost 0.10% per side

In-sample, a region of good results appears around a slow average of 75 days. Out of sample, results were better overall and the strongest region shifted towards shorter fast averages and a slow average of 50 to 75 days. Pairs with a long slow average of 150 to 200 days did worst out of sample. The broad shape is similar across periods, which supports the rank correlation result, but the best single cell moved. A user who fixed on one exact pair would have been lucky or unlucky depending on the period.

4.4 The effect of trading costs

Table 6. CAGR at different costs per side, Nifty 50, full sample

Rule	Round trips	0.00%	0.05%	0.10%	0.20%	0.30%
Buy and hold	1	10.9%	10.9%	10.9%	10.9%	10.9%
SMA 10/75	30	9.8%	9.6%	9.4%	8.9%	8.5%
SMA 50/200	10	3.8%	3.7%	3.7%	3.5%	3.4%
Price above SMA 200	63	5.3%	4.8%	4.4%	3.5%	2.6%
Price above SMA 20	189	9.0%	7.6%	6.2%	3.4%	0.8%
RSI 30/55	21	1.4%	1.3%	1.1%	0.8%	0.5%
Breakout 20/10	90	4.3%	3.6%	3.0%	1.7%	0.5%

Source: Author's calculations

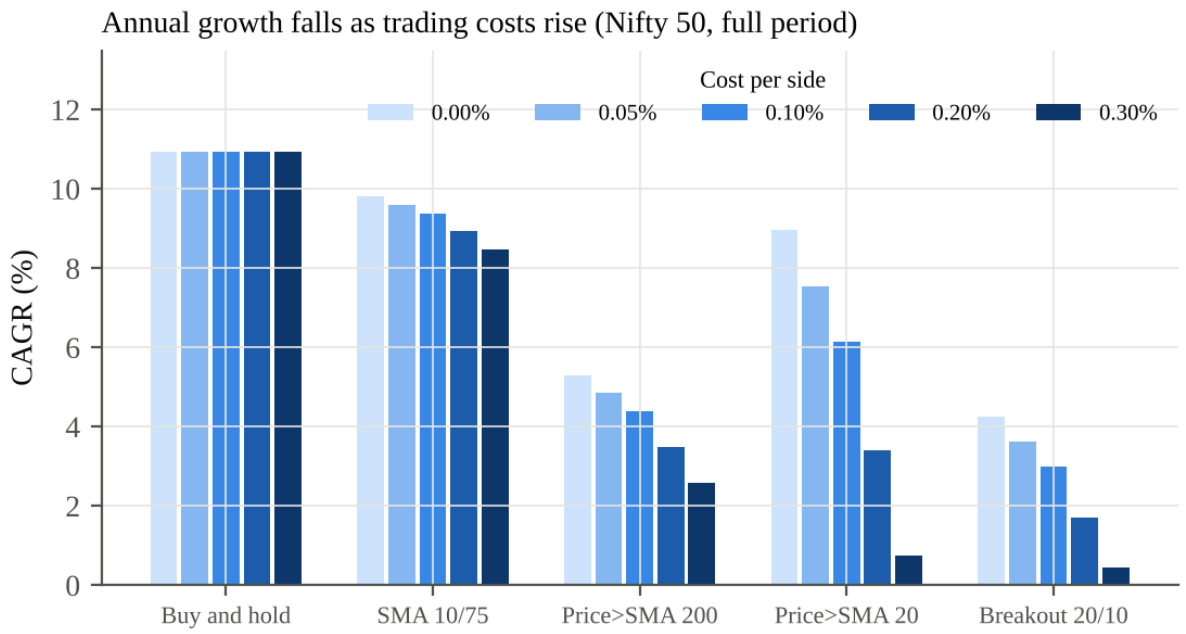


Figure 5. CAGR at different costs per side, Nifty 50

Source: Author's calculations

The effect of costs depends on how often a rule trades. The "price above SMA 20" rule made 189 round trips. With no costs it earned 9.0% a year, close to buy and hold, with a Sharpe ratio of 0.91. At 0.10% per side its CAGR fell to 6.2%, and at 0.30% it fell to 0.8%. The Sharpe ratio fell from 0.91 to 0.12. A backtest without costs would have made this rule look better than buy and hold on a risk-adjusted basis. With realistic costs it was clearly worse. By contrast, SMA 10/75, with 30 round trips, lost only about 1.3 percentage points of CAGR across the full range of costs.

4.5 Deflated Sharpe ratio and probability of backtest overfitting

Table 7. Formal measures of overfitting for the best in-sample rule

Measure	Nifty 50	Nifty Bank
Best in-sample rule	SMA 10/75	SMA 20/50
Annualised in-sample Sharpe ratio	0.87	1.08
Expected maximum Sharpe ratio from 74 trials (annualised)	0.61	0.75
Probabilistic Sharpe ratio (against zero)	0.99	1.00
Deflated Sharpe ratio	0.77	0.82
Probability of backtest overfitting (CSCV, 16 blocks, 2,000 splits)	0.36	0.35

Source: Author's calculations, following Bailey and López de Prado (2014) and Bailey et al. (2017)

Taken alone, each best rule looks highly significant: the probability that its true Sharpe ratio is above zero is 0.99 or more. But once the number of trials is allowed for, the picture changes. With 74 rules tried, the expected best Sharpe ratio from luck alone is 0.61 for the Nifty 50 and 0.75 for Nifty Bank. The deflated Sharpe ratios of 0.77 and 0.82 are below the usual 0.95 threshold, so neither best rule can be called a reliable discovery.

The estimated PBO of about 0.35 for both indices means that in roughly one split out of three, the rule that looked best on one half of the data ranked in the bottom half on the other. This is lower than the 0.5 that pure noise would give, so the rule set is not entirely random. But it is far too high for a user to trust the best in-sample rule without further checks.

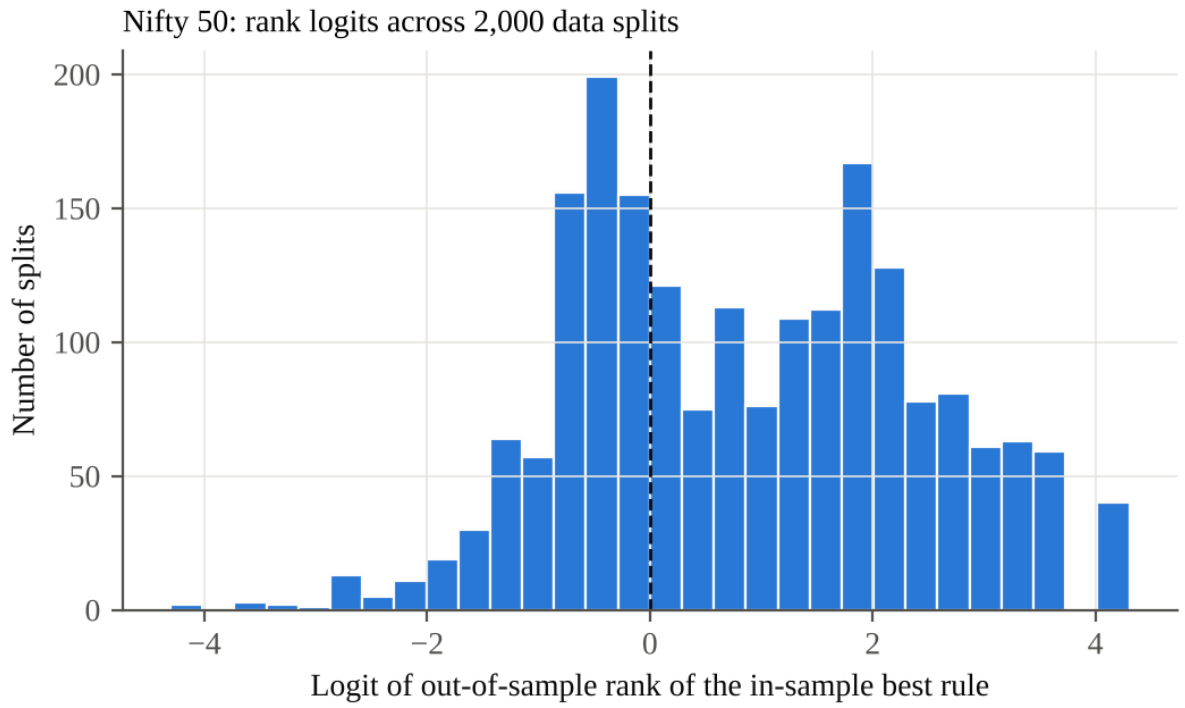


Figure 6. Distribution of rank logits across 2,000 data splits, Nifty 50

Source: Author's calculations. Values at or below zero mean the selected rule ranked at or below the median out of sample.

4.6 Random rules

Table 8. One thousand random rules on the Nifty 50, cost 0.10% per side

Measure	Value
Average in-sample Sharpe ratio	0.52
95th percentile in-sample Sharpe ratio	0.85
Best in-sample Sharpe ratio	1.17
Share beating buy and hold in-sample (Sharpe 0.78)	9.0%
Out-of-sample Sharpe ratio of the best in-sample random rule	0.29
Average out-of-sample Sharpe ratio of the top 10 in-sample random rules	0.45
Average out-of-sample Sharpe ratio of all random rules	0.51
Spearman correlation, in-sample vs out-of-sample	0.06

Source: Author's calculations

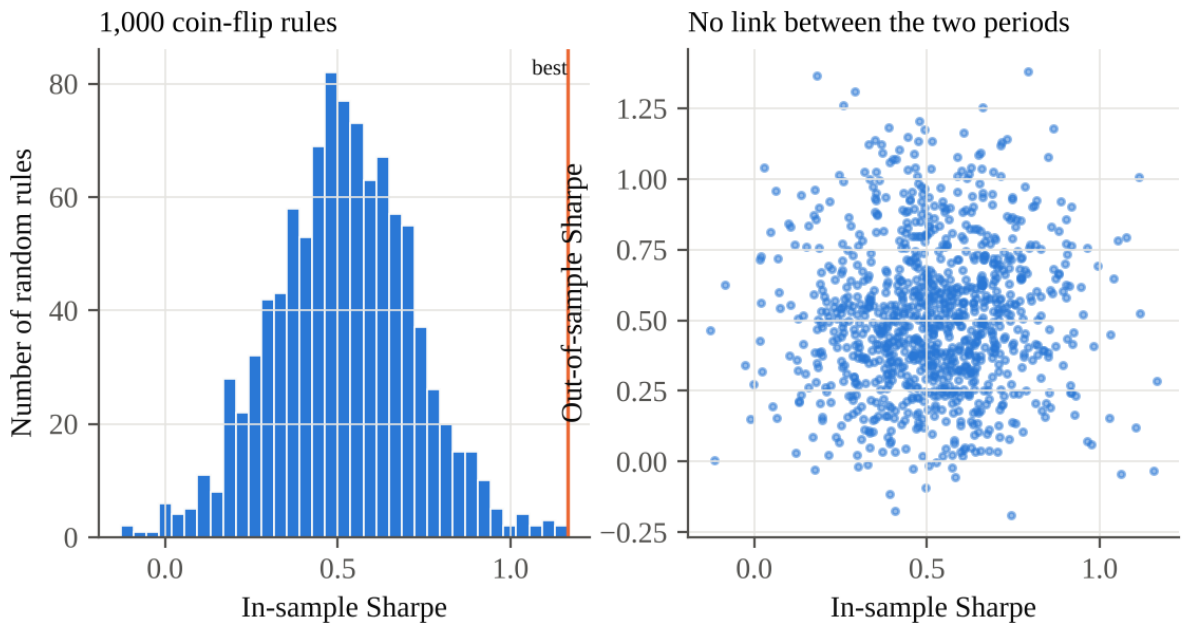


Figure 7. In-sample Sharpe ratios of 1,000 random rules and their link to out-of-sample results

Source: Author's calculations

The random rules make the danger of overfitting concrete. None of them has any logic, yet the best had an in-sample Sharpe ratio of 1.17, higher than every one of the 74 real rules and well above buy and hold. Ninety of the 1,000 beat buy and hold in-sample by chance. Out of sample, the best random rule fell to 0.29, and the top ten averaged 0.45, slightly below the average of all random rules. The rank correlation between periods was close to zero.

Two lessons follow. First, an in-sample Sharpe ratio of around 0.85 is within the normal range of luck for this index and exposure, so a real rule needs to clear a higher bar. Second, the real rules showed a much stronger link between periods (0.69 on the Nifty 50) than the random rules (0.06). This suggests that trend-type rules do capture something more than noise on this index, even if they do not beat buy and hold.

4.7 Walk-forward re-optimisation

Table 9. Walk-forward selection of moving average crossovers, Nifty 50, 2017 to 2026

Measure	Walk-forward rule	Buy and hold
CAGR	5.7%	12.3%
Volatility	10.7%	16.2%
Sharpe ratio	0.57	0.80
Maximum drawdown	-21.0%	-38.4%

Source: Author's calculations, cost 0.10% per side. Each year's rule is the crossover with the best Sharpe ratio over the previous five years.

The walk-forward test mimics a user who re-optimises every year. The selected pair changed often: SMA 40/150 in 2017 and 2018, SMA 50/150 in 2019, SMA 15/75 in 2020, SMA 30/50 in 2021, SMA 20/50 in 2022, SMA 5/50 in 2023 and 2024, SMA 5/75 in 2025 and SMA 15/75 in 2026. The result

was a CAGR of 5.7% against 12.3% for buy and hold, and a lower Sharpe ratio. Repeated selection of the recent winner did not add value. It reduced drawdown, but at a large cost in return.

5 Discussion

5.1 Answers to the research questions

1. 1. Do simple rules beat buy and hold after costs? On the evidence here, mostly no. On the Nifty 50, all timing rules earned less than buy and hold, and only one beat it on Sharpe ratio in-sample. Some rules did beat it on Sharpe ratio out of sample, largely by avoiding the 2020 crash. On Nifty Bank, very few rules beat buy and hold out of sample.
2. 2. Does the best past rule stay among the best? It depends on the index and the period. On the Nifty 50 the best rule held up; on Nifty Bank it did not. The formal measures (DSR below 0.95 and PBO about 0.35) say that a user should not rely on the best in-sample rule.
3. 3. How much do costs matter? A great deal for rules that trade often. A cost of 0.10% per side cut the CAGR of a 189-trade rule by almost three percentage points a year.
4. 4. What should a report show? This is discussed in Section 5.3.

5.2 Why trend rules reduce drawdowns but not returns

The trend and crossover rules reduced volatility and drawdowns by staying out of the market during long falls. The cost was missing part of the recoveries, since the rules re-entered only after prices had already risen. Indian indices rose strongly over the sample, so time out of the market was costly. These rules may suit an investor whose main concern is limiting deep losses, but they should be judged on that basis, not sold as a way to earn more than the market.

5.3 Lessons for no-code backtesting platforms

No-code and AI-based platforms make it easy to try many variations. The results above suggest that, without safeguards, users of such platforms are likely to choose rules that look good by chance. A platform that wants to help its users rather than mislead them could build the following into every report.

Table 10. Suggested standard features for retail backtest reports

Feature	Reason from this study
Default trading costs	A frequent-trading rule went from looking better than buy and hold to clearly worse once costs were added
Buy-and-hold benchmark shown alongside	All timing rules earned less than buy and hold on the Nifty 50
Automatic out-of-sample split	The best Nifty Bank rule lost two thirds of its Sharpe ratio out of sample
Count of variations tried	With 74 trials, a Sharpe ratio of 0.61 to 0.75 can be expected from luck alone
Deflated Sharpe ratio or a luck benchmark	Random rules reached a Sharpe ratio of 1.17 in-sample
Year-by-year returns	Drawdown benefits came mainly from one year, 2020
Parameter heatmaps	The best single parameter pair moved between periods

Source: Author

These features would not stop a determined user from overfitting. But they would make the risk visible at the point of decision, which is where it matters. For platforms aimed at students and first-time systematic traders, such as the one where this study was carried out, they also serve a teaching purpose.

5.4 Link to retail losses in India

SEBI's finding that costs pushed the share of loss-making F&O traders from 82.1% to 87.7% in FY26 matches the cost results in this paper on a smaller scale. Trading often is expensive, and small edges disappear. The study did not test options strategies, which account for most retail losses, but the same logic applies with more force, since options carry higher costs and wider spreads.

6 Limitations

- Price indices: the data excludes dividends, which understates buy and hold returns by roughly one to one and a half percentage points a year. Including dividends would make buy and hold look better still.
- No interest in cash: rules earned nothing while out of the market. With an interest rate of around 6% a year on idle cash, timing rules would improve, by roughly 2% a year for a rule out of the market a third of the time. This could change the ranking of some rules against buy and hold and should be tested in future work.
- One split: the main analysis uses one in-sample and one out-of-sample period. The PBO and walk-forward tests partly address this.
- Index-level rules only: the study does not test individual stocks, options or intraday rules, which are common on retail platforms.
- Simple cost model: a flat cost per side does not capture changes in spreads during stress, such as in March 2020.
- Limited rule set: 74 rules from four families is small compared with what a platform user might try, which means the overfitting problem in practice is likely to be larger than measured here.

7 Conclusion

This paper tested 74 simple trading rules on fourteen years of Nifty 50 and Nifty Bank data. On the Nifty 50, no rule beat buy and hold on return, and only one beat it on risk-adjusted return in the period used to choose rules. On Nifty Bank, the best rule from 2012 to 2019 lost most of its edge in 2020 to 2026 and fell behind buy and hold.

Formal checks agreed. The deflated Sharpe ratios were below the usual bar for a reliable result. The probability of backtest overfitting was about one in three. Random rules with no logic produced better in-sample results than any real rule. A user who re-picked the best recent rule each year earned less than half the return of buy and hold.

Costs mattered as much as the choice of rule. Rules that traded often lost most of their gains once costs were included.

None of this means backtesting is useless. It means a backtest is only as honest as the report around it. Platforms that make testing easy should also make the risks easy to see. Out-of-sample results,

costs, a benchmark and a count of trials should be standard. For India's young and growing base of retail investors, that could be the difference between a tool that teaches and a tool that misleads.

References

[1] Bailey DH, Borwein JM, López de Prado M, Zhu QJ (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.

[2] Bailey DH, Borwein JM, López de Prado M, Zhu QJ (2017). The probability of backtest overfitting. *Journal of Computational Finance*, 20(4), 39–69.

[3] Bailey DH, López de Prado M (2014). The deflated Sharpe ratio: correcting for selection bias, backtest overfitting and non-normality. *Journal of Portfolio Management*, 40(5), 94–107.

[4] Barber BM, Lee YT, Liu YJ, Odean T (2014). The cross-section of speculator skill: evidence from day trading. *Journal of Financial Markets*, 18, 1–24.

[5] Barber BM, Odean T (2000). Trading is hazardous to your wealth: the common stock investment performance of individual investors. *Journal of Finance*, 55(2), 773–806.

[6] Brock W, Lakonishok J, LeBaron B (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47(5), 1731–1764.

[7] Davis FD (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.

[8] Faber MT (2007). A quantitative approach to tactical asset allocation. *Journal of Wealth Management*, 9(4), 69–79.

[9] Fama EF (1970). Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, 25(2), 383–417.

[10] Gunasekarage A, Power DM (2001). The profitability of moving average trading rules in South Asian stock markets. *Emerging Markets Review*, 2(1), 17–33.

[11] Harvey CR, Liu Y (2015). Backtesting. *Journal of Portfolio Management*, 42(1), 13–28.

[12] Harvey CR, Liu Y, Zhu H (2016). ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5–68.

[13] Lo AW (2004). The adaptive markets hypothesis. *Journal of Portfolio Management*, 30(5), 15–29.

[14] López de Prado M (2018). Advances in financial machine learning. *1st edition*. Wiley.

[15] Mitra SK (2011). How rewarding is technical analysis in the Indian stock market? *Quantitative Finance*, 11(2), 287–297.

[16] Moskowitz TJ, Ooi YH, Pedersen LH (2012). Time series momentum. *Journal of Financial Economics*, 104(2), 228–250.

[17] National Stock Exchange of India (2026). Investor accounts cross the milestone of 26 crore (260 million). *Press release*, 7 June 2026. https://nsearchives.nseindia.com.

[18] Park CH, Irwin SH (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, 21(4), 786–826.

[19] Securities, Exchange Board of India (2025). Safer participation of retail investors in algorithmic trading. *SEBI circular dated* 4 February 2025. https://www.sebi.gov.in.

[20] Securities, Exchange Board of India (2026). Studies on profitability of individual traders in the equity derivatives segment, FY26. *Press Release No*. 50/2026, 20 August 2026. https://www.sebi.gov.in.

[21] Sullivan R, Timmermann A, White H (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, 54(5), 1647–1691.

[22] Thadikaran B (2026). eod2_data: end of day historical data for NSE stocks and indices. *GitHub*. https://github.com/BennyThadikaran/eod2_data.

[23] White H (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097–1126.