MPM MARKETS RESEARCH · WORKING PAPER

Fair Value Gaps Work — Until You Try to Trade Them

A four-market investigation of whether the most-marketed concept in retail trading has a tradeable edge — and why its apparent edge does not survive honest execution.

Research Dataset: S&P 500 (ES), Nasdaq-100 (NQ), Gold (GC), and Silver (SI) Futures · 1-minute base data resampled to 5-minute, 15-minute and 1-hour, April 2019 – May 2026

Author: Dhaval Barot · Published by MPM Markets – Market Probability Model · mpmmarkets.com

Abstract

We test the widely-taught belief that a Fair Value Gap (FVG) — a three-candle price imbalance — marks a level where price will reliably react, using seven years of data across four liquid futures markets, three timeframes, and roughly 40,000 detected occurrences. We measure not only whether price reacts at these levels, but whether it does so more than a random level at the same distance — the only test that separates a genuine signal from noise. We find the reaction is real as a measurement (approximately five percentage points above random) but carries **no tradeable edge**. Across five independent trade constructions — including the 5- and 15-minute versions most heavily marketed on social media — the effect is either a coin flip or, at its best, exactly consumed by realistic trading costs. The one construction that appears strongly profitable proves to be an artifact: it depends on resolving trade exits on coarse hourly bars, which smuggles in look-ahead bias; measured honestly on 1-minute data, the apparent edge collapses from a profit factor of ~2.4 to ~1.0. The result is consistent across all four instruments and every timeframe. We report the descriptive finding and the negative tradability result together, with full sample sizes and methodology, so that other researchers can verify, extend, or refute this work.

1. Motivation

The Fair Value Gap — a three-candle pattern leaving an "imbalance" that price is said to return to and react from — is among the most widely taught setups in modern retail trading. It dominates social-media trading education, almost always presented as a reliable level to trade against. Yet that confidence rests on remarkably little evidence: presentations show hand-picked successful examples and rarely address how often it works, whether it beats a random level, or whether the fast timeframes it is sold on actually perform.

We set out to measure the behaviour directly and — separately — to test honestly whether any measured tendency translates into an exploitable edge. We treat these as two different questions, because a statistically real pattern and a profitable strategy are not the same thing. MPM Markets operates on a principle of *reaction, not prediction*: we test whether a setup carries genuine, capturable forward information before treating it as tradeable.

Research Snapshot

Research question — Can a Fair Value Gap be traded profitably?

Markets tested — 4 futures markets Years of data — 7 Occurrences analysed — ~40,000

Timeframes — 5-minute, 15-minute, 1-hour Constructions tested — 5

Main result — **Reaction is real; no tradeable edge after honest costs**

2. Data and Definitions

Instruments. Four CME-group futures across three asset classes: S&P 500 (ES) and Nasdaq-100 (NQ) equity indices; gold (GC) and silver (SI) metals. Chosen so any genuine effect would replicate across unrelated markets, while a market-specific fluke would stand out as isolated.

Sample. 20 April 2019 through 24 May 2026. Base data is 1-minute resolution (~2.5 million bars per instrument), resampled to 5-minute, 15-minute, and 1-hour. Data collected and normalized by the author; timezone handling verified per file.

Pattern definition (pre-registered). Bullish FVG: candle-1 high below candle-3 low. Bearish: candle-1 low above candle-3 high. The zone spans that gap, with three reference levels — near edge (first touched on return), midpoint (consequent encroachment), and far edge (full-fill / invalidation). A displacement variant additionally requires the middle candle's range to exceed 1.5× the median of the prior 20 bars. Definitions were fixed before measurement.

Significance and honest execution. Every reaction rate is tested against a matched-random benchmark: a level at the same distance and direction from price, over the same forward window. Crucially, every stop-and-target outcome is resolved on 1-minute data, counting stop or target strikes only from the minute after entry actually fills. This single discipline — established after it overturned a preliminary positive result (Section 4) — is what separates this study from most FVG backtests.

3. Principal Finding: The Reaction Is Real but Faint

Before any trading rule, we measured whether price reacts at FVG midpoints more than at matched-random levels. Across all three index/metal markets, on hourly, 4-hour, and daily data, the FVG reaction rate exceeded the random rate in 34 of 36 tested cells, by a median of roughly five percentage points. The two exceptions were small-sample daily cells whose uncertainty exceeded the effect. The tendency is consistent in sign across markets and directions.

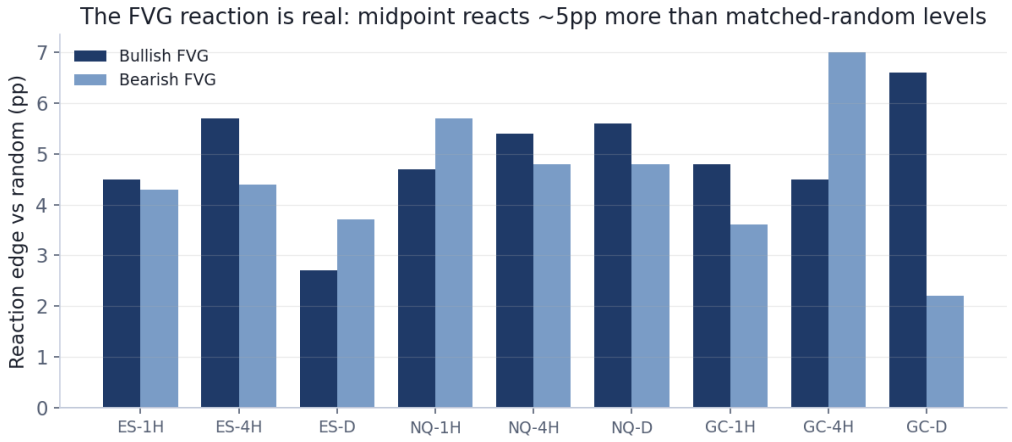

Figure 1. FVG midpoint reaction edge over matched-random levels under the raw definition, by market and timeframe. All cells are positive with a median of roughly five percentage points, confirming the reaction is real as a measurement.

Taken naively this looks promising. But a level that reacts, say, 45% of the time versus 40% for a random level is real only in the sense that most of the time price still does not react favourably. Whether that thin tilt survives stops, targets, and costs is a separate question — and the rest of this paper answers it.

4. The Look-Ahead Illusion

One construction appeared overwhelming. Entering at the midpoint with a volatility-scaled stop and a fixed target, and resolving exits on 1-hour bars, produced win rates near 73% and profit factors of 2.2 to 2.5 — a spectacular, apparently profitable result across all three markets.

> **The central distinction of this paper**
>
> **This is an artifact, not an edge.** When a single 1-hour bar spans both the stop and the target, coarse-bar resolution cannot know which was struck first, and a naive backtest credits the favourable one — or credits a target reached before

the entry would even have filled within that hour. Saying "these trades hit target" is close to saying "trades that worked, worked." It is mechanically true and untradeable, because the coarse bar hides the true order of events. Resolving the identical trades on 1-minute data — the only sequence a trader could actually experience — collapses the result.

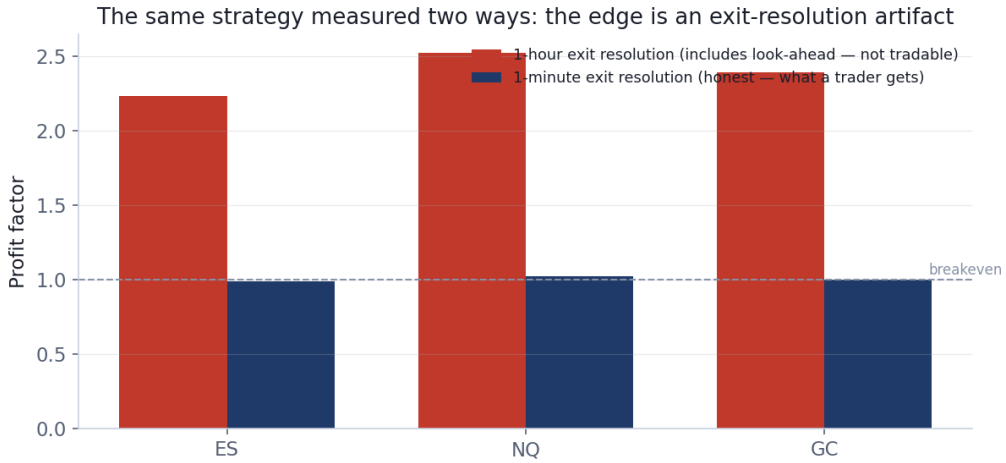

Figure 2. The identical strategy resolved two ways. Coarse 1-hour exit resolution includes look-ahead bias and reports a profit factor near 2.4 in every market, while honest 1-minute resolution collapses it to approximately 1.0, showing the apparent edge to be an exit-resolution artifact.

Win rate fell from roughly 73% to roughly 50%. The entire apparent edge was intrabar look-ahead. This is not a subtle academic point: it is the single most likely reason a retail FVG backtest shows a profit. **Any stop-and-target strategy must be resolved on data fine enough to establish the true sequence of events within the bar.**

5. Tradability Testing: Five Constructions, All Fail

With exits resolved honestly on 1-minute data, we tested five distinct ways to trade the FVG. Each was chosen to give the concept a fair chance; several were designed specifically to address why the previous one failed.

#	Construction	Result
1	Midpoint entry, ATR stop, fixed target	Apparent edge is look-ahead; honest PF ≈ 1.0 (Section 4)
2	Enter on close through near edge (breakout)	No edge; win rate = chance at every payoff
3	Near-edge bounce, 14-hour ATR stop, R:R sweep	No edge; win rate on the breakeven line at 1:3, 2:3, 1:1
4	Near-edge, 14-day ATR stop / target	Tiny apparent edge is drift; drawdown 3–4× profit
5	Wick + volume reversal at midpoint	Real +3 to +6 pp win-rate edge on 1H — fully consumed by costs

5.1 The coin-flip fingerprint (constructions 2 and 3)

Entering on the near edge and sweeping the risk-reward ratio, the win rate landed exactly on the breakeven line every time — about 25% at 1:3, 40% at 2:3, 50% at 1:1 — across all four instruments including silver. That is the mathematical fingerprint of an entry with no predictive information: the outcome matches pure chance for whatever payoff geometry is chosen.

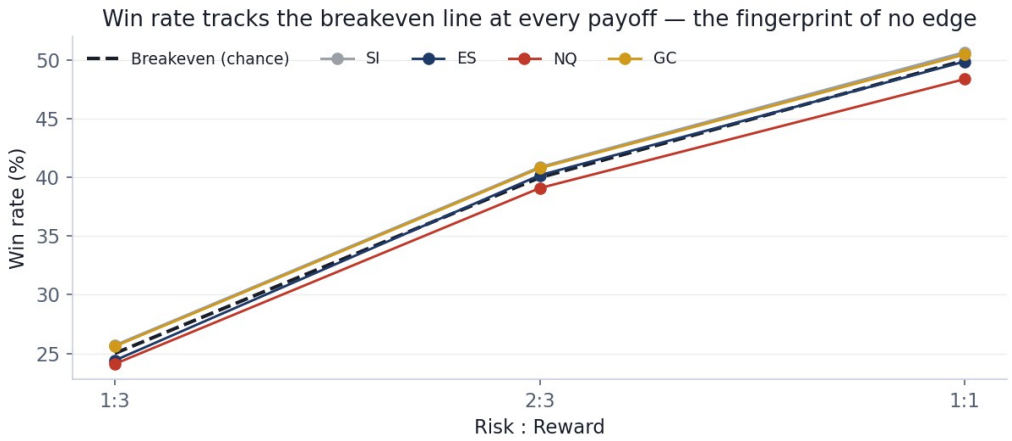

Figure 3. Near-edge entry win rate versus risk-reward ratio across all four markets. Every instrument tracks the breakeven line implied by the payoff geometry, and silver behaves no differently from the indices or gold, the mathematical fingerprint of an entry carrying no information.

5.2 The daily-ATR mirage (construction 4)

Widening stops to a 14-day ATR produced a small apparent edge of about three points — but only at the 1:3 setting, vanishing to zero at 1:1, and accompanied by drawdowns three to four times larger than total profit. The tell that this is drift, not an FVG effect: gold, the instrument with the least directional drift, was the weakest and lost money, while the higher-drift index markets showed the small positive. A genuine entry edge appears at every payoff; a drift artifact appears only where wide stops and distant targets let the market's long-run direction accumulate.

5.3 The best honest case (construction 5)

The strongest construction waits at the midpoint for a rejection candle — a wick against the level, closing on the reaction side, with above-average volume — then enters on the next bar with a stop below the far edge. This filter genuinely works in the narrow sense that matters: on the 1-hour chart it lifted the win rate a real three to six percentage points above breakeven, consistently across ES, NQ, and GC. For the first time, a confirmed entry beat chance.

But detectable is not tradeable. Across eighteen configurations (three markets × three timeframes × two targets), exactly one finished positive — gold on 1-hour at a 3R target, earning roughly six dollars per trade over seven years, indistinguishable from zero and negative under any additional slippage. At a 1:1 target on 1-hour, the three markets net to approximately flat. The filter finds the tilt; costs erase it.

6. The Timeframe Finding: The Marketed Version Is the Worst

The most practically useful result concerns timeframe. FVG scalping is overwhelmingly marketed on the 5- and 15-minute charts. On honest testing, these are the worst-performing timeframes, without exception. The edge improves as the timeframe lengthens.

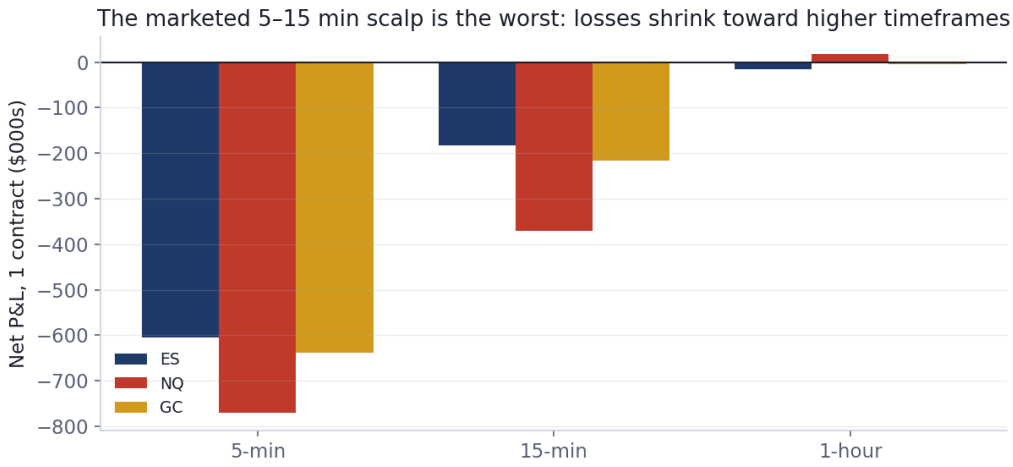

Figure 4. Net profit and loss of the best-filtered rule (wick-and-volume reversal, one-to-one target) by timeframe, one contract. Losses are deepest on the 5-minute chart, shrink on the 15-minute, and approach breakeven only on the 1-hour, showing the heavily marketed fast-scalp to be the least viable version.

The mechanism is straightforward: on fast timeframes the gaps are tiny relative to noise, the stop is small and easily struck, trade counts explode into the tens of thousands, and commission plus slippage overwhelm any faint edge. A wick-and-volume "reversal" on a 5-minute bar is largely random. The concept sold as a fast-timeframe scalp is, on evidence, the single least viable expression of an already-unviable idea.

7. Conclusions

Key findings at a glance

- ✓ Tested ~40,000 FVGs across four markets · five constructions · three timeframes
- ✓ Every outcome resolved on 1-minute data and benchmarked against random levels
- ✗ No tradeable edge at any entry, any timeframe, after honest costs
- ✗ The marketed 5–15 minute scalp is the worst-performing version, not the best
- o The one profitable-looking result proved to be a 1-hour exit-resolution artifact

The Fair Value Gap is descriptive, not tradeable. Across four futures markets and roughly 40,000 occurrences, price reacts at FVG levels slightly more than at random — a genuine, measurable effect — but that tilt does not forecast price strongly enough to trade profitably at any entry, on any timeframe, once execution is honest. At its best, a volume-confirmed reversal on the 1-hour chart sits exactly at breakeven. The reaction is real; the edge is not.

This distinction is central to the Market Probability Model: a measurable market tendency is valuable research, but it is not automatically a tradeable edge. The gap between the two is precisely where honest testing earns its keep.

8. Practical Takeaway

For a practitioner, the useful conclusion is informational, not directional. The FVG is a legitimate way to read where the market left an imbalance — price does interact with these levels slightly more than chance — but it is not, by itself, a reason to enter a trade. Do not trade a setup because it is popular; verify every belief statistically; and be especially wary of backtests that resolve exits on coarse bars, because that is where a phantom edge hides. Anyone trading FVGs profitably is most likely being helped by some other factor they have not isolated.

9. Limitations

Bull-market regime. 2019–2026 rose overall. Behaviour in a sustained bear market is untested; the null is strong but not a universal law.

Modelled execution. Costs and slippage are modelled, not live; the realistic-fill and trade-through stress tests bound this, but only live trading is definitive.

Mechanical interpretation. Each construction tests the FVG claim as commonly stated and taught. Discretionary, context-aware applications are deliberately outside scope; the null applies to the mechanical claim.

Finite construction space. Five constructions across three timeframes is broad but not exhaustive. Our negative result applies to what was tested.

10. What a Researcher Could Try Next

A conditioning variable — session, higher-timeframe context, or liquidity state — that meaningfully raises the reaction rate above the cost line, rather than the raw reaction alone. A sustained bear-market regime, the one condition this study could not test. Alternative gap definitions or adaptive volatility-scaled tolerances. Each should be pre-registered and held to the same 1-minute exit-resolution and random-benchmark discipline before any weight is placed on it.

> **Why we publish a study with no edge**
>
> Most published trading research highlights strategies that succeeded. MPM also publishes carefully executed null findings, because knowing which widely held ideas do *not* survive rigorous testing is equally valuable. Eliminating false positives reduces survivorship bias, guards against overconfidence, and helps researchers focus their time on hypotheses with genuine potential. We would rather show our work — including where it did not lead to a tradable strategy — than present a flattering result we could not stand behind. This is the essence of reaction, not prediction: we do not assume a setup works because it is popular; we test it, report what we find, and trust only what survives.

11. Selected References

This work is empirical and self-contained, but it sits within an established tradition of evidence-based scepticism toward technical trading claims. The following place it in that context.

Aronson, D. R. (2006). Evidence-Based Technical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals. Wiley.

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

Bailey, D. H., & Lopez 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.

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

Lo, A. W., & MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies, 3(3), 431-467.

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. Journal of Finance, 55(4), 1705-1765.

Pardo, R. (2008). The Evaluation and Optimization of Trading Strategies (2nd ed.). Wiley.

White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097-1126.

12. Reproducibility

Parameter	Value
FVG definition	c1.high < c3.low (bull) / c1.low > c3.high (bear)

Parameter	Value
Displacement variant	middle-candle range > 1.5 × median range, prior 20 bars
ATR length	14 (hourly or daily, per construction)
Entry levels tested	near edge, midpoint; entry-at-close and next-bar-open
Stops tested	far edge, 1.0–2.0 × ATR (14-hour and 14-day)
Targets tested	1R, 2R, 3R, and time-only exit
Exit resolution	1-minute bars, strikes counted only after entry fill
Reversal filter	wick ≥ 0.5 × range on reaction side + above-average volume
Timeframes	5-minute, 15-minute, 1-hour
Cost model	commission + slippage per round turn, per instrument
Base data	~2.5M 1-minute bars per instrument, 2019–2026

About the research. This study was conducted and authored by Dhaval Barot and published by MPM Markets – Market Probability Model. Questions, replication requests, and research discussions may be submitted through the contact form at mpmmarkets.com/contact.

Suggested citation. *Barot, D. (2026). Fair Value Gaps Work — Until You Try to Trade Them: A Four-Market Investigation of Whether the Most-Marketed Concept in Retail Trading Has a Tradeable Edge. MPM Markets.*