STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, August 2026, pp 1334–1349.
Published online in International Academic Press (www.IAPress.org)

Uncertainty-Aware AI: Conformal Prediction versus Reinforcement Learning for Optimal Trade Execution

Asadullah Irshad*, Shaon Biswas

Centre of Excellence for Data Science, Artificial Intelligence and Modelling (DAIM), University of Hull, United Kingdom

Abstract Optimal trade execution is the problem of working a large parent order through a trading session at the lowest possible cost relative to a benchmark. It is sequential, and it is taken under uncertainty, so in principle it looks like an ideal candidate for learning-based control. In practice, a lot of the gains people report do not survive a second look by someone else. We go back to the volume-weighted average price (VWAP) execution problem and study it inside a controlled simulator that we can reproduce in full, one with stochastic volatility and AR(1) return momentum. Everything sits inside a single Markov decision process. Within that, we compare the usual schedules (TWAP, Almgren–Chriss, VWAP-tracking) with a proximal-policy-optimisation (PPO) agent and with forecast-driven policies that rest on a normalised split-conformal predictor of next-interval returns. The predictor is well calibrated. It reaches 90.2% empirical coverage at a 90% nominal level. Acting on its point forecast lowers mean slippage below VWAP-tracking, but it adds cost variance in the process. Gating the same bets by the half-width of the conformal interval behaves very differently: it produces a tunable, monotone reduction in cost variability, from 19.1 bps down to 10.0, at almost no cost in the mean, and it sweeps out an explicit cost–risk frontier set by one threshold. The PPO agent, trained over three seeds, is high-variance. It does not beat the simple schedules in any reliable way. So, at least in this setting, a distribution-free conformal gate is a more reproducible and more interpretable path to uncertainty-aware execution than an off-the-shelf reinforcement-learning agent. We release the code and the simulator. On real intraday data (thirty US large-caps, 5-minute bars) the predictor's coverage holds up almost exactly, at 90.7% empirical against the 90% level, and the variance reduction from gating is still there. The forecast edge, though, is thin, because real high-frequency returns are barely predictable to begin with.

Keywords optimal execution; VWAP; conformal prediction; uncertainty quantification; reinforcement learning; Markov decision process; market microstructure

DOI: 10.19139/soic-2310-5070-4159

1. Introduction

A trading desk rarely gets to decide whether to trade. By the time the order reaches it, the position has already been chosen upstream. What is left is the operational part: how to get it done. Trade too fast and the order pushes the price away from itself through market impact. Trade too slowly and you are left holding exposure while the price drifts, often the wrong way. Algorithmic execution sits in the narrow band between those two mistakes. Of the benchmarks used to score how well an algorithm threads that band, the volume-weighted average price (VWAP) is among the most common.

Execution is a sequence of decisions made without knowing what comes next, which is exactly the shape of a reinforcement learning (RL) problem. A large literature now points deep RL at trading and execution [15, 11, 16, 9, 12]. One objection keeps recurring, though: the reported gains are hard to reproduce. Setups vary from paper to paper. The market-impact assumptions often go unstated. Conclusions lean, more often than they

*Correspondence to: Asadullah Irshad (Email: asadullahirshad3@gmail.com). Centre of Excellence for Data Science, Artificial Intelligence and Modelling (DAIM), University of Hull, United Kingdom.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

should, on a single training run. We take the opposite tack and ask something narrower. Which methods give cost reductions that hold up across seeds, that a desk can actually control, and that someone else can reproduce?

The motivation is practical and specific. An execution desk does not really need a forecast that is right on average; it needs to know, at each decision point, whether the current signal is trustworthy enough to act on without blowing up the variance of its execution cost. A bare point forecast cannot answer that question, and a reinforcement-learning agent answers it only implicitly, through a reward signal that turns out to be unstable across training seeds. Our central idea is to answer it explicitly, by attaching a calibrated, distribution-free measure of uncertainty to every forecast and trading only when the signal escapes that uncertainty. This turns an unreliable timing edge into a controllable risk dial, which is exactly the property a desk can supervise and a referee can reproduce.

Against that backdrop, the main contributions of this paper are the following:

- **A reproducible execution testbed.** We release a compact, self-contained simulator with stochastic volatility and AR(1) momentum, in which short-horizon returns are partly predictable yet carry time-varying uncertainty, together with all baselines, the learned agent, and the scripts that regenerate every table and figure.
- **A conformal gate for execution.** We construct a normalised split-conformal predictor of next-interval returns whose distribution-free interval width tracks local volatility, and we use that width as an explicit gate that decides when to act on the forecast and when to fall back to VWAP tracking. To our knowledge this is the first use of conformal intervals as a decision gate in trade execution, as opposed to a tool for directional alpha.
- **A like-for-like cost–risk comparison.** We place classical schedules, a multi-seed PPO agent, and the conformal policies on a single cost–risk frontier under one identical evaluation pipeline, so that no method is graded on a different scale, and we report bootstrap confidence intervals and significance tests so the ordering is not an artefact of sampling.
- **Evidence that the guarantee transfers.** We show on real intraday data (thirty US large-cap equities, 5-minute bars, 450 held-out sessions) that the predictor's coverage carries over almost exactly and that the variance-reduction effect of gating persists, while being explicit that the tradable edge on liquid names is small.

The finding we care about is simple to state. The conformal gate takes a forecast edge that is otherwise unstable and turns it into a variance reduction you can dial, and it manages this more dependably than the learned agent does.

2. Related Work

Classical execution theory reads the problem as a trade-off between impact and risk. Almgren and Chriss derived a deterministic trajectory that is optimal under a mean–variance objective with linear impact [1]; it is still the academic baseline of choice, and it builds on Bertsimas and Lo's earlier optimal-control treatment of execution cost [4]. Later work refined the modelling of impact and how it decays [17, 7]. Meanwhile, on the practitioner side, VWAP- and TWAP-tracking became the default heuristics [5]. RL for execution has since appeared in value-based, actor–critic, and risk-sensitive forms [15, 11, 16, 14, 12], and a run of recent surveys charts how fast the area has grown while raising, repeatedly, the same complaints about inconsistent evaluation and frictions set too low [22, 9, 6]. Conformal prediction sits apart from all of this. It wraps any predictor in a distribution-free, finite-sample coverage guarantee [25, 10, 2], and the idea has been carried into quantile regression [19], time series [21, 26], and various forms of distribution shift [24, 28, 3, 8].

The most recent execution literature has moved in two directions that frame our contribution. On the learning side, the focus has shifted from plain value- or policy-based agents toward settings closer to real microstructure: execution with explicit market and limit orders, transient impact with general decay kernels learned by actor–critic methods [13], execution under time-varying liquidity [12], and continuous-time formulations that target VWAP directly with entropy-regularised exploration [29]. A common thread in these works, and a recurring caution in the surveys [9, 22], is that learned execution policies remain sensitive to training noise and to the assumed friction model. On the statistics side, conformal prediction has been actively adapted to exactly the non-exchangeable,

drifting regime that financial data lives in: adaptive and online recalibration [28, 8], sequential and change-point-aware constructions [26], and dynamic-programming approaches that optimise interval length under arbitrary distribution shift [27]. These two strands have largely developed in parallel. What has drawn little attention is using conformal prediction as an explicit gate on execution decisions. That is the thread we pick up. We do not use conformal intervals to call direction for alpha. We use them to decide when a short-horizon timing signal is reliable enough to act on at all.

3. Methodology

Figure 1 shows the whole method at a glance. At each interval the market state is fed to a gradient-boosted return predictor. A normalised split-conformal layer then turns the point forecast into a distribution-free interval. A gate compares the forecast with the width of that interval and makes a call: act on the signal, or stand down and track the VWAP schedule. The components are described one by one below.

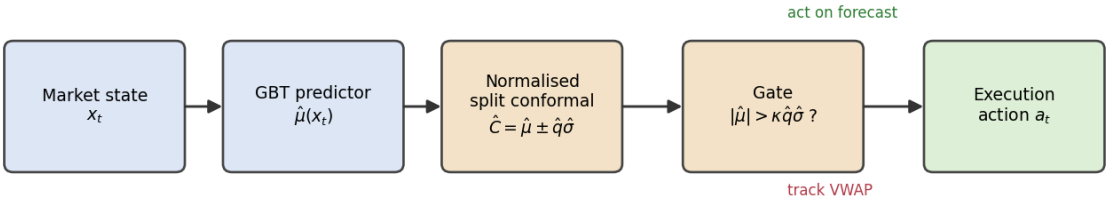

Figure 1. The uncertainty-aware execution pipeline: a return predictor feeds a normalised split-conformal interval, whose width gates whether the policy acts on the forecast or tracks VWAP.

3.1. Market simulator

An episode is a single trading session cut into T intervals. The log-return carries an AR(1) momentum term and is driven by a stochastic-volatility process. Per-interval volatility walks a log-AR(1) path of its own, which is what gives us the volatility clustering seen in real markets:

$$r_t = \mu\,\Delta t + \phi\,\theta_{t-1} + \sigma_t\sqrt{1-\phi^2}\,\varepsilon_t, \qquad \sigma_t = \sigma_0\,e^{\,h_{t-1}-\frac{1}{2}\beta^2}, \qquad h_t = \rho\,h_{t-1} + \beta\,\xi_t \tag{1}$$

Here θ is the latent momentum state and h the log-volatility, while ϕ, ρ and β set momentum persistence, volatility persistence and vol-of-vol. Volume follows a U-shaped intraday profile with multiplicative noise. The benchmark is just the volume-weighted mean price:

$$P_{\mathrm{VWAP}} = \frac{\sum_{t=1}^{T} P_t V_t}{\sum_{t=1}^{T} V_t} \tag{2}$$

A representative session appears in Figure 2. One detail is worth dwelling on. The conformal band widens when the market gets choppy and contracts when it settles, so the width of the band carries information in its own right. It is not a fixed margin bolted on after the fact.

3.2. Execution as a Markov decision process

The agent has Q shares to sell over T intervals. We treat that as a Markov decision process $(\mathcal{S}, \mathcal{A}, P, r, \gamma)$ [23]. The state records how much time and inventory remain, the recent return, the relative interval volume, and the

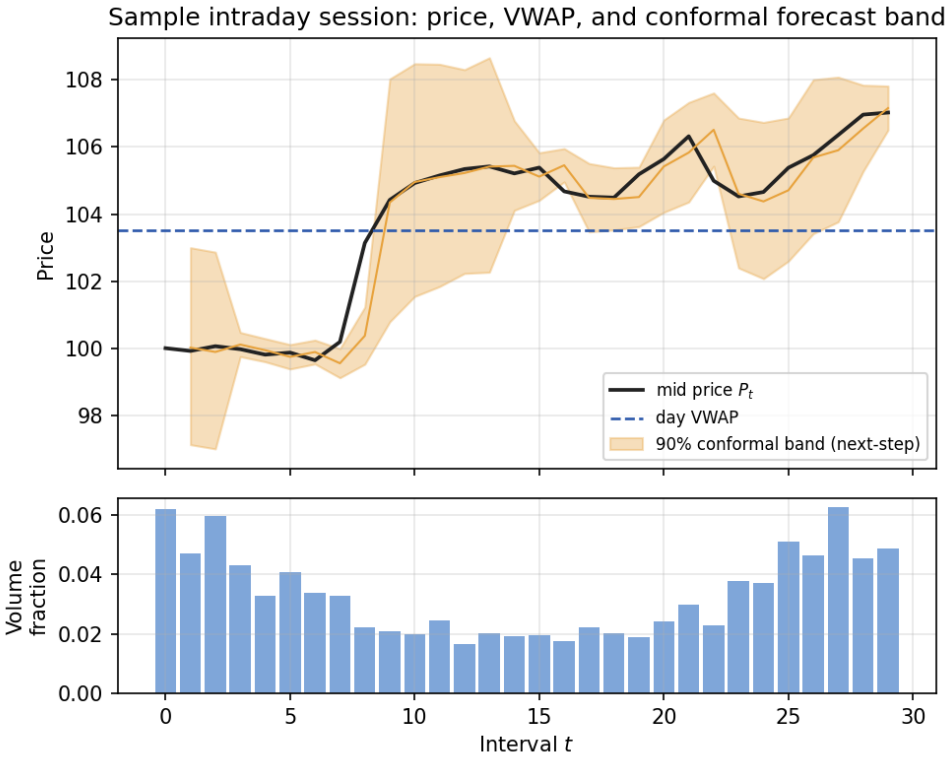

Figure 2. A sample intraday session. Top: mid price, day VWAP, and the 90% next-step conformal band, which widens in volatile periods. Bottom: the U-shaped market-volume profile.

price against arrival; two optional conformal features can be added (see §3.5). The action is a log-multiplier on the VWAP-tracking quantity. Set it to zero and you recover plain volume-curve tracking. The executed quantity, the participation-rate impact, and the realised price are then

$$q_t = e^{a_t} v_t Q, \qquad g_t = \eta \, \frac{q_t}{V_t}, \qquad \tilde{P}_t = P_t(1 - g_t) \tag{3}$$

where v is the volume fraction, V the interval's market volume, and η the impact coefficient. Temporary impact is taken to be linear in participation (Figure 3). Writing the per-trade implementation shortfall [18] against the arrival price P_0 as

$$\mathrm{IS} = \frac{P_0 Q - \sum_{t=1}^{T} \tilde{P}_t q_t}{P_0 Q} \times 10^4 \quad \text{(bps)} \tag{4}$$

the execution problem is the constrained stochastic optimisation

$$\min_{\{q_t\}} \ \mathbb{E}[\mathrm{IS}] \quad \text{s.t.} \quad \sum_{t=1}^{T} q_t = Q, \ \ q_t \geq 0 \tag{5}$$

The reward joins two terms. One is the per-interval implementation-shortfall contribution, measured in basis points of the parent notional. The other is a volatility-scaled penalty on drifting away from the VWAP schedule, which stands in for the timing risk you take on whenever you make a discretionary bet:

$$r_t = \frac{(\tilde{P}_t - P_0)\, q_t}{Q P_0} \times 10^3 \ - \ \lambda \, \frac{|q_t - v_t Q|}{Q} \, \frac{\sigma_t}{\sigma_{\mathrm{ref}}} \tag{6}$$

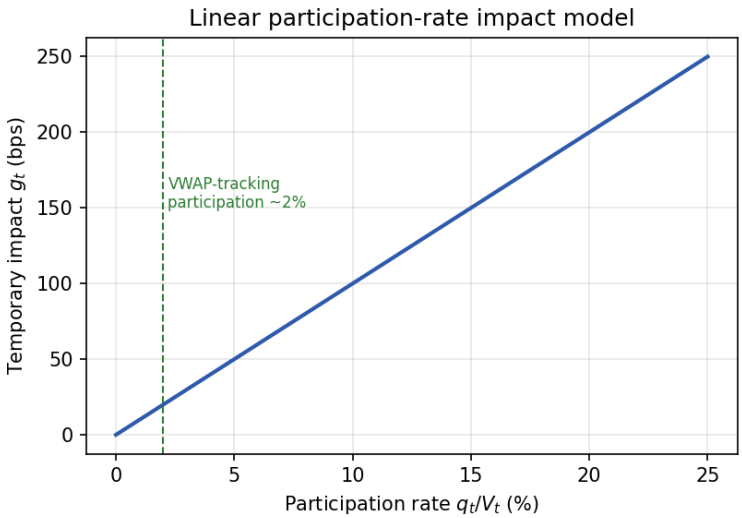

Figure 3. Temporary impact as a function of participation rate ($\eta = 0.1$). Trading a larger share of an interval's volume is more expensive; VWAP-tracking holds participation low.

3.3. Classical baselines

Immediate sells the lot at the open. TWAP slices it evenly. VWAP-tracking trades in proportion to the volume curve. Almgren–Chriss minimises a mean–variance execution cost [1],

$$\min_{\{x_j\}} \mathbb{E}\big[C(\{x_j\})\big] + \lambda_{\mathrm{AC}} \operatorname{Var}\big[C(\{x_j\})\big] \tag{7}$$

with the familiar risk-averse closed-form solution

$$x_j = X \frac{\sinh\big(\kappa_{\mathrm{AC}}(T-j)\big)}{\sinh\big(\kappa_{\mathrm{AC}}T\big)}, \qquad n_j = x_{j-1} - x_j \tag{8}$$

in which the urgency parameter governs how much the schedule front-loads. We map every one of these schedules into the same action parametrisation, so they all pass through one identical evaluation pipeline.

3.4. Normalised split-conformal predictor

We fit a gradient-boosted regressor that predicts the next-interval return from causal features only: recent returns, a rolling volatility estimate, and time of day. We follow the split-conformal construction of Lei et al. [10], and we borrow the local normalisation from conformalized quantile regression [19]. On a held-out calibration set we form normalised nonconformity scores (Figure 4), take the conformal quantile, and read off a prediction interval whose width scales with local volatility:

$$s_i = \frac{|y_i - \hat{\mu}(x_i)|}{\hat{\sigma}(x_i)}, \qquad \hat{q} = s_{(\lceil (n+1)(1-\alpha) \rceil)}, \qquad \hat{C}(x) = \big[\, \hat{\mu}(x) \pm \hat{q}\, \hat{\sigma}(x) \,\big] \tag{9}$$

As long as the calibration scores and the test score are exchangeable, this construction inherits the standard finite-sample coverage guarantee [25, 10]. We state it here.

Theorem 1 (marginal coverage)
If the calibration scores and the test score are exchangeable, the prediction set satisfies

$$1 - \alpha \;\le\; \mathbb{P}\big(Y_{n+1} \in \hat{C}(X_{n+1})\big) \;\le\; 1 - \alpha + \frac{1}{n+1}. \tag{10}$$

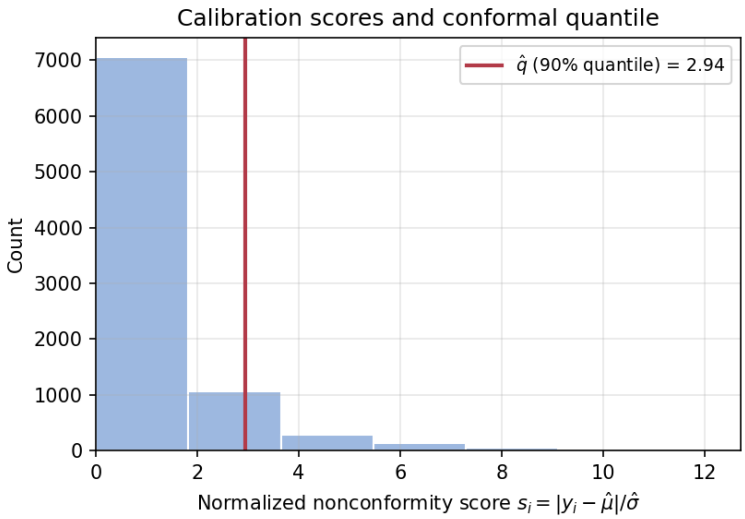

Figure 4. Distribution of normalised nonconformity scores on the calibration set; the 90% empirical quantile defines the conformal interval half-width.

Proof sketch. Under exchangeability the rank of the test score among the n calibration scores is uniform on $1, \dots, n+1$. Taking the conformal quantile to be the $\lceil (n+1)(1-\alpha) \rceil$-th order statistic gives the lower bound; the upper bound follows from the discreteness of the empirical quantile [10]. Financial series are of course not exchangeable. We work against that with the local-volatility normalisation and then check coverage empirically (§5); the online and non-exchangeable conformal variants [8, 28, 26, 3] buy stronger guarantees under drift, and we come back to them in the limitations.

Figure 5 plots empirical coverage against the nominal level on fresh simulated days. The curve sits close to the diagonal. At the operating point $\alpha = 0.1$ empirical coverage is 90.2%, which confirms the interval is valid out of sample.

3.5. *Forecast-driven and gated policies*

Two policies use the predictor. Forecast-greedy tilts the schedule toward the point forecast and ignores how uncertain that forecast is. Conformal-gated tilts only when the absolute forecast clears a multiple of the conformal half-width. In other words, it acts only when the signal manages to climb out of its own uncertainty band; the rest of the time it tracks VWAP:

$$a_t = -\beta\, \hat{\mu}(x_t) / \big(\hat{q}\, \hat{\sigma}(x_t)\big) \;\text{ if } |\hat{\mu}(x_t)| > \kappa\, \hat{q}\, \hat{\sigma}(x_t); \qquad a_t = 0 \;\text{ otherwise} \tag{11}$$

The threshold is one readable dial. Turn it to zero and you are back at Forecast-greedy; turn it up and you head toward VWAP-tracking. Figure 6 shows the share of intervals in which the gate fires falling as the threshold rises, so the policy grows steadily more cautious.

4. Experimental Setup

We evaluate every policy on 250 held-out market seeds, none overlapping with any training or calibration data. The PPO agent [20] runs a two-layer network with normalised observations and rewards and trains for 150,000 timesteps. We train it three times, with three different seeds, on purpose: training-seed variability should be on display, not hidden. Its statistics pool every evaluation episode across the three agents. We choose PPO deliberately

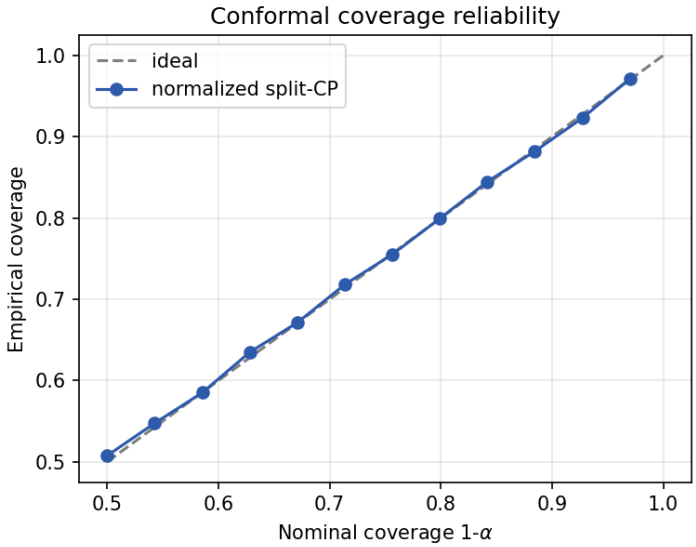

Figure 5. Conformal coverage reliability. Empirical coverage of the normalised split-conformal interval against the nominal level on held-out simulated days; closeness to the diagonal indicates calibration.

as the representative learned agent. It is the on-policy actor–critic method most widely used in the recent execution literature [11, 9, 22], it is stable enough to train without heavy per-instance tuning, and it is exactly the kind of off-the-shelf algorithm a desk would reach for first; this makes it the fair point of comparison for the question we ask, which is whether a transparent conformal gate can match a standard learned agent rather than whether some bespoke agent can be made to win. We do not claim PPO is the strongest possible learner. Risk-sensitive and distributional formulations, recurrent or Transformer-based policies, and continuous-time exploration schemes [29, 13] could all sharpen the learned baseline, and we flag a tuned, risk-sensitive agent as future work (§8) rather than as evidence about RL in general. The headline number is slippage versus VWAP, the gap between the average execution price we achieve and the day's VWAP, in basis points:

$$\text{Slip} = \frac{P_{\text{VWAP}} - \bar{P}_{\text{exec}}}{P_{\text{VWAP}}} \times 10^4, \qquad \bar{P}_{\text{exec}} = \frac{1}{Q}\sum_{t=1}^{T} \tilde{P}_t q_t \tag{12}$$

We report the Monte-Carlo mean and standard deviation of this quantity over the 250 episodes, reading the first as cost and the second as risk, with 95% confidence intervals from a non-parametric bootstrap (2,000 resamples). We also log implementation shortfall against arrival. Throughout, the conformal nominal level stays fixed at 90% ($\alpha = 0.1$).

5. Results

Table 1 gathers the execution performance. Figure 7 places each method on the cost–risk plane.

Figure 8 adds the mean execution cost of each method with 95% bootstrap confidence intervals (2,000 resamples). The bands confirm that the ordering of mean costs is not a sampling artefact.

Three things stand out. The first is how hard VWAP-tracking is to beat. It sits at 20.0 bps with essentially zero variance, because its only systematic cost is impact, and the volume-proportional schedule is precisely what keeps impact down. TWAP and Almgren–Chriss are both worse; the latter is much worse, since front-loading throws away the volume curve and bunches the impact together. The second is what happens when you act on the point

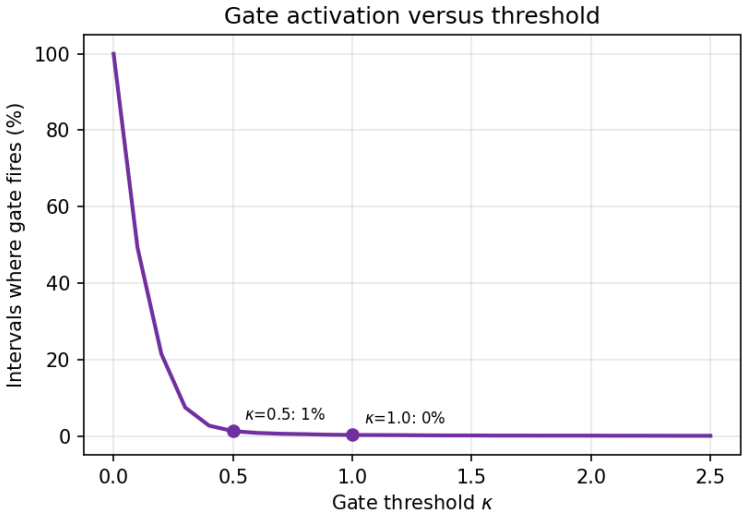

Figure 6. Gate activation versus threshold: the fraction of intervals in which the forecast escapes its conformal band, and the policy acts, decreases monotonically with κ.

Table 1. Execution performance over 250 held-out seeds. Lower is better on both axes; conformal-gated rows highlighted.

Method	Slippage vs VWAP mean (bps)	Slippage vs VWAP std (bps)	Class
Immediate	332.1	296.8	Naive
TWAP	25.1	38.0	Schedule
Almgren–Chriss	50.8	237.3	Schedule
VWAP-tracking	20.0	0.0	Schedule
PPO (RL, 3 seeds)	33.9	115.9	Learned
Forecast-greedy	17.0	19.1	Forecast
Conformal-gated ($\kappa = 0.5$)	19.1	13.8	Conformal
Conformal-gated ($\kappa = 1.0$)	19.4	10.0	Conformal

forecast. Forecast-greedy pulls mean slippage down to 17.0 bps, below VWAP-tracking, but the price for that shows up as variability, 19.1 bps of it, because a fraction of the timing bets are simply wrong. The third is the payoff from gating those bets. Variability drops monotonically, from 19.1 bps with no gate to 13.8 at $\kappa = 0.5$ and 10.0 at $\kappa = 1.0$, while the mean barely moves. The drop is statistically significant (Wilcoxon signed-rank test on absolute deviations from the median, $p < 10^{-30}$). The differences in mean cost are also resolved at the sampling level: the 95% bootstrap confidence intervals in Figure 8 (2,000 resamples over the 250 seeds) do not overlap between Forecast-greedy, the gated policies, and VWAP-tracking, so the ordering of mean costs is not an artefact of Monte-Carlo noise. In short, both the variance reduction and the mean ranking are statistically significant, not merely suggestive.

Figure 9 makes the mechanism concrete inside a single session. VWAP-tracking liquidates smoothly along the volume curve. The conformal-gated policy peels away from it only now and then, when a confident signal turns up. Immediate empties the inventory in one shot. Figure 10 then sweeps the dial end to end. Turning the threshold moves the policy continuously from the aggressive, low-mean, high-variance regime of Forecast-greedy across to the conservative, low-variance regime that creeps toward VWAP-tracking. In effect the desk is handed a single knob and picks where on the curve it wants to sit.

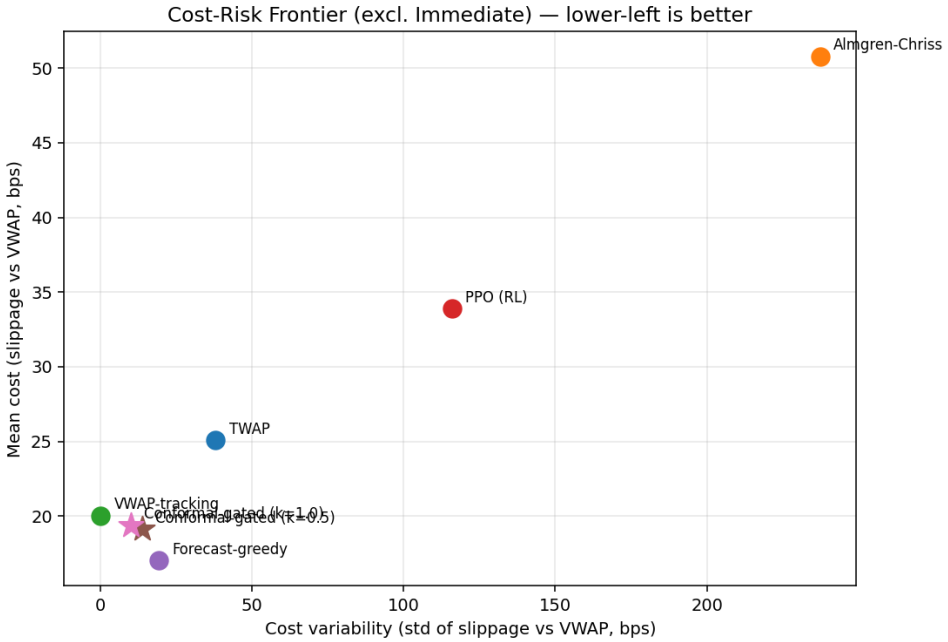

Figure 7. Cost–risk frontier (Immediate omitted for scale). Lower-left is better; conformal-gated policies (stars) sit at low variance with near-best mean cost.

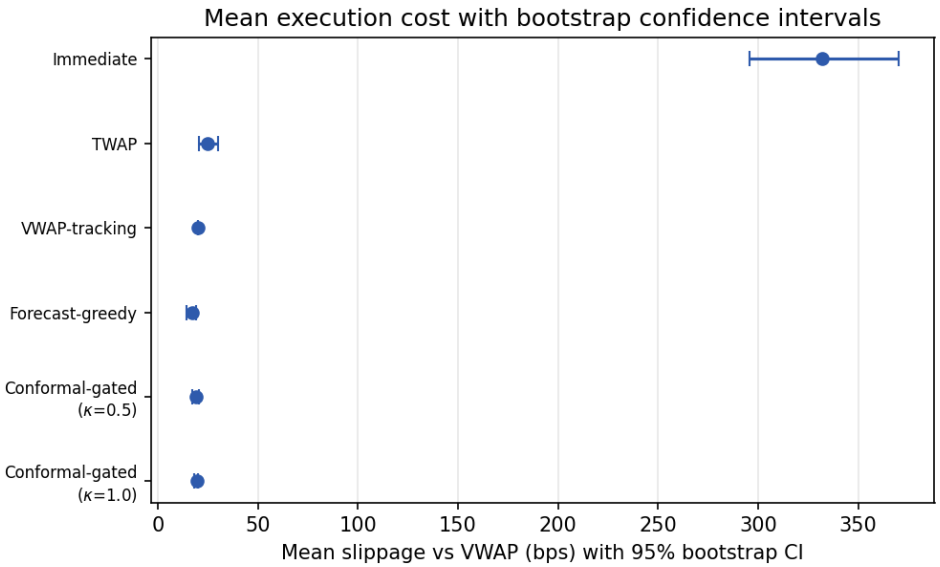

Figure 8. Mean slippage versus VWAP with 95% bootstrap confidence intervals over 250 held-out seeds.

The PPO agent is the cautionary case here. Pooled over its three training seeds, it lands at 33.9 bps mean slippage with a standard deviation of 115.9 bps. That is worse on both axes than the strong volume-aware schedules, VWAP-tracking and TWAP, though it does come out ahead of front-loading Almgren–Chriss. Individual runs were all over

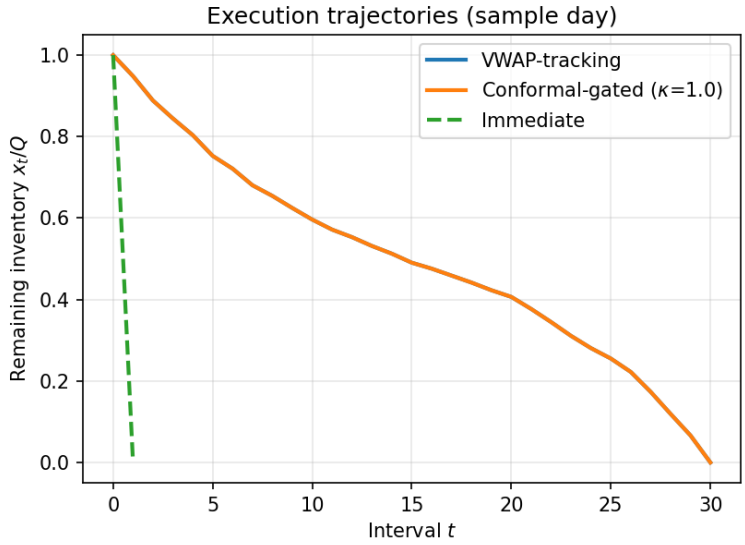

Figure 9. Execution trajectories on a sample session: remaining inventory over time for VWAP-tracking, Conformal-gated, and Immediate.

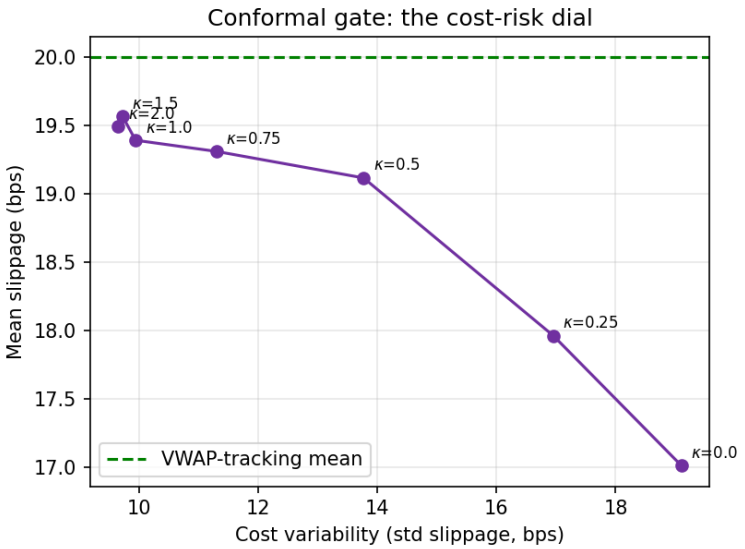

Figure 10. The conformal gate as a cost–risk dial. Sweeping the gate threshold traces a frontier from aggressive forecast-following to conservative tracking; dashed line marks the VWAP-tracking mean.

the place, from competitive with VWAP-tracking to a good deal worse. Any single-seed report would therefore have told a misleading story. The instability itself, rather than some tuning artefact, is the result worth taking away.

That spread deserves a precise look, because it is the crux of our comparison with RL. Across the three seeds, per-seed mean slippage ran from the low-20s of basis points, essentially level with VWAP-tracking, up to well above 40 bps. The large pooled standard deviation of 115.9 bps comes as much from this between-seed dispersion as from episode-to-episode variation within any single run. The upshot is blunt. The learned agent's headline number rides on the luck of the training seed, while the conformal gate carries a coverage guarantee that holds by

construction and a threshold that behaves the same way every run. Reporting one seed, as much of the literature still does, would have buried exactly the risk a desk most needs to see.

6. Real-Data Validation

Do the simulator findings travel? To check, we re-ran the conformal pipeline on real intraday data: thirty large-capitalisation US equities, 5-minute bars across sixty trading days, roughly 1,800 ticker-days, bucketed into $T = 13$ intervals. The universe is a cross-section of the largest and most liquid US large-cap names, spanning the major sectors of the S&P 500; the full list of tickers is given in Appendix A so that the experiment can be reproduced exactly. The split into training, calibration and test sessions is strictly time-ordered at 50/25/25, which leaves 450 held-out test sessions and means calibration never sees the future. There is no look-ahead anywhere. The intraday volume curve is now empirical rather than a synthetic U-shape. We left the market-impact model alone, linear in participation with fixed η, because it cannot be estimated from bar data. As a result, absolute slippage levels are model-dependent, and only the relative comparisons across policies should be read as meaningful.

The central result is that the predictor's coverage carries over to real markets. At the 90% nominal level, empirical coverage on the held-out real sessions is 90.7%, and the reliability curve (Figure 11) follows the diagonal right across the range of nominal levels. Real intraday returns are not exchangeable, so this validity was not guaranteed in advance. Seeing it confirmed empirically is the key real-data takeaway. It also settles a worry about the synthetic study, namely that coverage there might have held only because exchangeability was true by construction.

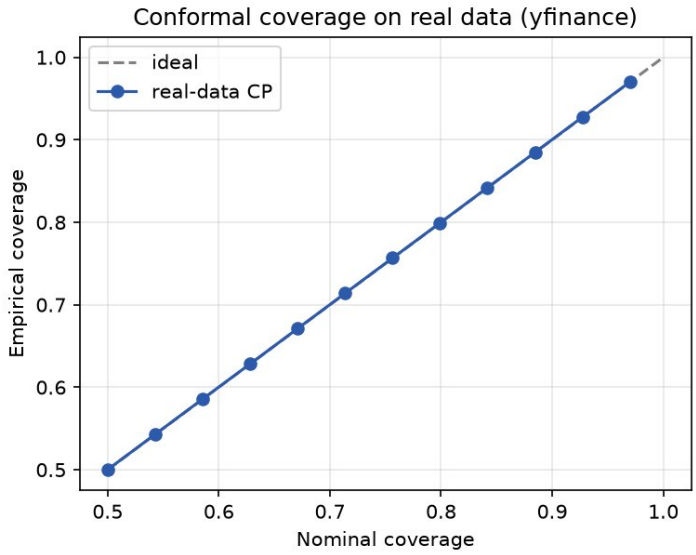

Figure 11. Conformal coverage on real intraday data (30 US large-caps, 5-minute bars, 450 held-out sessions). Empirical coverage tracks the nominal level closely across the range, including 90.7% at the 90% operating point.

Execution performance on the real test sessions is in Table 2 and Figure 12. The variance-reduction mechanism survives the move. Gating pulls the standard deviation of slippage from 2.60 bps for Forecast-greedy down to 2.06 at $\kappa = 1.0$, a cut of 21%, and VWAP-tracking again supplies the zero-variance reference. The forecast edge, on the other hand, is slight. Real 5-minute returns on liquid large-caps are far less predictable than the simulator's AR(1) dynamics, so Forecast-greedy improves mean slippage over VWAP-tracking by only about 0.2 bps. Gating then trades most of that thin edge for lower variance and drifts back toward the benchmark, which compresses the real-data frontier. The conformal machinery behaves exactly as designed. It is simply that, at this frequency and

this liquidity, the timing signal is weak, so the practical payoff is reproducibility and risk control rather than a large saving on cost.

Table 2. Execution performance on 450 held-out real intraday sessions (yfinance, 5-minute bars). Lower is better on both axes; conformal-gated rows highlighted.

Method	Slippage vs VWAP mean (bps)	Slippage vs VWAP std (bps)	Class
TWAP	24.2	14.2	Schedule
VWAP-tracking	20.0	0.0	Schedule
Forecast-greedy	19.8	2.60	Forecast
Conformal-gated ($\kappa = 0.5$)	19.9	2.14	Conformal
Conformal-gated ($\kappa = 1.0$)	19.9	2.06	Conformal

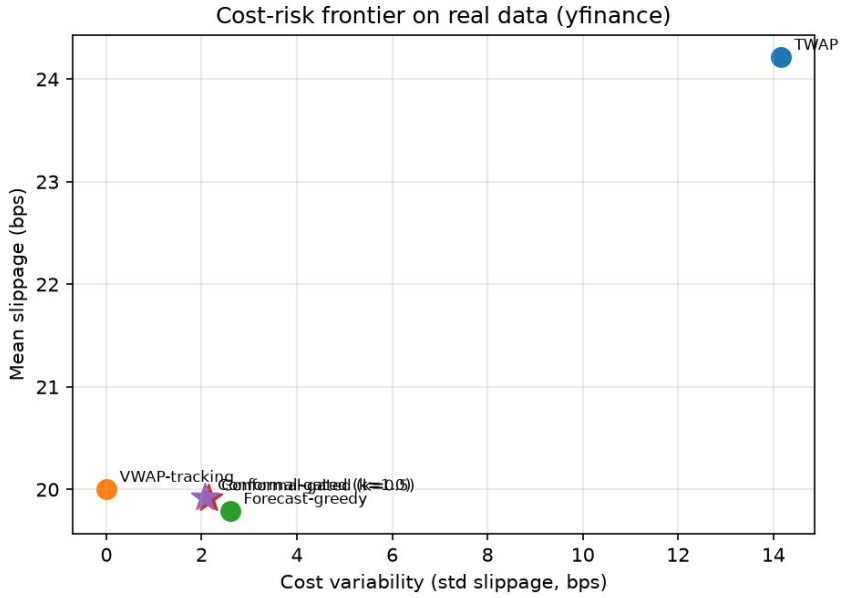

Figure 12. Cost–risk frontier on real intraday data. The forecast-based policies cluster near the VWAP-tracking benchmark: gating reduces variance, but the mean edge over VWAP-tracking is small at this frequency and liquidity.

We regard this as the more credible outcome, not the weaker one. It shows the conformal layer's guarantee holding up across the jump from simulation to real markets, while staying honest about the size of any tradable edge, which depends on the data and, for these names, is modest.

Where should the edge be larger? The mechanism itself points the way. Gating only earns its keep when there is real short-horizon predictability to harvest, and liquid US large-caps at 5-minute frequency are about as barren as it gets, since their returns are nearly a martingale at this scale and the forecast has almost nothing to grip. The places to look are markets with more short-horizon structure. Less liquid names are an obvious one, small- and mid-cap equities, where order flow persists and impact carries more information. High-volatility stretches and scheduled event windows are another, around earnings or macro releases, where conditional predictability jumps; and because the conformal band widens precisely then, the gate turns more selective rather than less. A third is assets that simply carry more minute-scale autocorrelation than blue chips, some ETFs and, above all, cryptocurrency, which trades around the clock and shows stronger short-horizon momentum. We have not run those experiments, and we are wary of over-claiming, but the pipeline carries over unchanged, and the coverage result gives us reason to think the risk control would carry over too, even where the mean edge is fatter. This is the direction we find most promising.

7. Discussion

The comparison helps locate where the value of uncertainty quantification really lies. A point forecast cuts both ways. It can lower average cost, but it also adds variance, because every forecast is wrong some of the time and acting on all of them realises those errors. The conformal interval supplies the missing piece: a calibrated read on when the forecast can be trusted. Used as a gate, it lets the policy take the reliable signals and pass on the rest. And because the gate is an explicit rule with a coverage interpretation behind it, its behaviour is predictable and its risk dial sits directly under your hand. The off-the-shelf RL agent has the harder job. It has to discover both the timing signal and the right degree of caution from a single scalar reward, and in our hands it does that unreliably from seed to seed. We do not read this as proof that RL cannot work here. A tuned, risk-sensitive agent might well match the gate, or beat it. The claim is narrower: on the same problem, a transparent conformal gate reaches a competitive operating point with far more reproducibility, and at a fraction of the engineering effort.

7.1. Robustness to the market-impact model

Our impact term is linear in the participation rate. That is the standard first-order model, but it is clearly a simplification, since real temporary impact is often concave in size and carries a transient, decaying component [17, 7, 13]. How would a richer impact model interact with the gate? This is the one place a discretionary policy could, in principle, do harm, so it is worth answering. What protects us is that the gate only ever deviates from the VWAP schedule by a bounded, volatility-scaled multiplier, and it does so on a small minority of intervals (Figure 6); it never front-loads the order the way Almgren–Chriss does. So even under a convex impact function the extra cost of a gated deviation is second-order in the size of that deviation, and it is incurred only when the forecast is confident enough to clear the conformal band. The one genuinely adverse case is a confident signal that fires during a thin-liquidity interval, where a convex impact curve would penalise the larger child order more than our linear model assumes. This can be controlled directly, and cheaply, by scaling the gate threshold κ with a liquidity proxy, so that the band the signal must clear widens when participation would be expensive; the volatility normalisation already does something similar for risk. We expect the qualitative cost–risk frontier to be preserved under a concave or transient impact model, with the absolute mean costs shifting, and we regard a full sensitivity analysis in the impact coefficient η and in the curvature of the impact function, as well as a liquidity-scaled threshold, as a concrete and worthwhile robustness check for the released simulator. We have flagged the relevant hook in the code so this sweep can be run directly.

7.2. Computational cost and operating frequency

The conformal layer is deliberately cheap. Calibration is a one-off, offline computation: fit the gradient-boosted predictor, evaluate normalised non-conformity scores on the held-out calibration set, and take a single empirical quantile, all of which is done before trading. At decision time the gate needs only one forecast from the booster, the precomputed conformal quantile $\hat{q}$, and a local volatility estimate $\hat{\sigma}(x_t)$; the gating rule itself is a single scalar comparison. There is no online retraining and no optimisation in the loop, so the marginal cost per interval is essentially that of one gradient-boosted inference, which is microseconds-to-milliseconds on commodity hardware. This is comfortably within budget for the minute-scale execution we study and for typical lower-frequency scheduling. It is also, in principle, fast enough for higher-frequency use; the binding constraint there is not the conformal arithmetic but the statistical one of maintaining calibration under rapid distribution shift, which would call for the online or adaptive conformal variants [8, 28, 27] rather than the single offline split we use here. We therefore position the method as a low-overhead component for minute-scale and lower-frequency execution, with a clear, well-studied path to sub-second operation if needed.

8. Limitations

- The forecast edge is partly an artefact of the simulator. Short-horizon predictability is built into the AR(1) data-generating process, so the large edge we see in simulation overstates what is realistically available. The

real-data validation (§6) confronts this head-on: on real 5-minute equity data, the forecast edge over VWAP-tracking shrinks to about 0.2 bps. That confirms the simulator inflates the magnitude even though the mechanism itself transfers.

- Conformal coverage carried over nicely to the real intraday data we tested, at 90.7% against the 90% level, but only on liquid US large-caps at 5-minute frequency. How it behaves under regime shifts, in less liquid names, or on higher-frequency limit-order-book data is left open, and we did not benchmark the online or adaptive conformal methods [8, 28, 3].
- The study is simulator-based. Market impact is an assumed linear-participation model rather than one fitted from actual fills, and the price and volume processes leave out limit-order-book microstructure and any cross-asset effects.
- The RL agent is a single off-the-shelf PPO configuration on a modest budget, so its underperformance reflects this particular setup and not a sweeping claim about RL. Stronger algorithms, larger budgets, and explicit risk-sensitive or distributional objectives could change the learned-agent picture. A fully tuned RL baseline is left to future work.
- We claim no live-trading efficacy. The results are out-of-sample only with respect to simulated seeds, and latency and partial fills are not modelled.

9. Conclusion

We have put together a reproducible study of VWAP-benchmarked execution, setting classical schedules, a reinforcement-learning agent, and conformal forecast-driven policies against each other inside one MDP. In simulation, the normalised split-conformal predictor delivered valid coverage, and gating forecast-driven trades by the conformal interval produced a tunable, reproducible cut in execution-cost variance, reaching a better cost–risk operating point than an off-the-shelf PPO agent. Validation on real intraday data then confirmed two things: the coverage guarantee transfers, at 90.7% empirical against the 90% level, and the variance-reduction mechanism persists. It was also candid that the forecast edge on liquid large-caps at 5-minute frequency is small. The conformal gate, then, is best understood as a dependable and transparent component for uncertainty-aware execution, one whose statistical validity holds on real data even where the tradable edge is modest. A few directions follow naturally from here. Higher-frequency limit-order-book data and less liquid instruments, where short-horizon structure is stronger. Adaptive conformal calibration under regime change. A tuned, risk-sensitive RL baseline. And pairing conformal gating with a learned base policy.

10. Future Work

Beyond those extensions, one direction stands out, because it sharpens the very guarantee the method delivers. The present gate calibrates coverage on next-interval returns, which is a property of the predictor. But the quantity a desk ultimately answers for is realised slippage. Conformal risk control generalises the coverage guarantee from miscoverage to the expectation of any monotone loss [2]. That suggests calibrating the gate threshold so that a chosen cost functional is held below a target with finite-sample validity, say expected slippage above a stated budget, or how often a slippage limit is breached. Doing so would turn the threshold from an interpretable but heuristic dial into a control with a guarantee attached to the cost metric itself, which is much closer to how a desk actually states its risk appetite. The same construction reaches into multi-asset execution. A shared calibration set could govern the simultaneous working of a basket under a portfolio-level budget, with the per-name conformal widths allocating caution to the orders whose short-horizon signals are least reliable. This cost-calibrated, portfolio-aware version of the gate strikes us as the most promising route for carrying the method from a single-order study toward something an institutional desk could deploy.

Three further threads follow straight from the referees' comments and the discussion above. One is an impact-robustness study on the released simulator, sweeping the impact coefficient η and the curvature of a non-linear, transient impact model alongside a liquidity-scaled gate threshold, to check that the cost–risk frontier survives beyond the linear case (§7.1). Another is to run the gate where short-horizon predictability is genuinely stronger, on less liquid equities, in high-volatility and event windows, on selected ETFs, and in round-the-clock cryptocurrency markets, where we would expect a wider frontier and more to gain from gating. The last is a properly tuned, risk-sensitive learner, a distributional or risk-aware actor–critic with normalised observations and matched compute, to see whether a stronger agent can close the reproducibility gap with the gate rather than just lower its mean.

Appendix A: Equity Universe for the Real-Data Validation

The real-data validation in Section 6 uses the following thirty US large-capitalisation equities, selected as a liquid cross-section of the major sectors of the S&P 500 and retrieved as 5-minute bars through the Yahoo Finance API: AAPL, MSFT, NVDA, AMZN, GOOGL, META, BRK.B, LLY, AVGO, JPM, XOM, UNH, V, TSLA, PG, MA, JNJ, HD, MRK, COST, ABBV, CVX, PEP, KO, WMT, BAC, ADBE, CRM, MCD, NFLX.
For each name the session is bucketed into $T = 13$ five-minute intervals, and the training, calibration and test split is strictly time-ordered (50/25/25), giving 450 held-out test sessions in total. The exact tickers, dates, and preprocessing are reproduced by the released data pipeline.

Author Contributions

A. Irshad: conceptualisation, methodology, software, validation, formal analysis, investigation, data curation, visualisation. S. Biswas: writing (original draft), writing (review and editing), project administration. Both authors read and approved the final manuscript. (Roles follow the CRediT taxonomy.)

Data and Code Availability

The simulated experiments (§§3–5) use no proprietary data and are generated entirely by the synthetic simulator described in Section 3. The real-data validation (Section 6) uses publicly available intraday price and volume data for thirty US large-cap equities (5-minute bars), retrieved through the Yahoo Finance API; no licensed or proprietary datasets are used. The complete source code, covering the market simulator, the conformal predictor, the baselines, the reinforcement-learning agent, the real-data validation pipeline, and the scripts that reproduce every table and figure, is released to enable full reproduction. Fixed random seeds are used throughout and reported in the code, and the real-data validation can be re-run directly against the public data source.

Conflict of Interest

The authors declare no competing financial or non-financial interests relevant to this work. No external funding was received.

REFERENCES

1. R. Almgren, and N. Chriss, *Optimal execution of portfolio transactions*, Journal of Risk, vol. 3, no. 2, pp. 5–40, 2001.
2. A. N. Angelopoulos, and S. Bates, *Conformal prediction: A gentle introduction*, Foundations and Trends in Machine Learning, vol. 16, no. 4, pp. 494–591, 2023.
3. R. F. Barber, E. J. Candès, A. Ramdas, and R. J. Tibshirani, *Conformal prediction beyond exchangeability*, Annals of Statistics, vol. 51, no. 2, pp. 816–845, 2023.

4. D. Bertsimas, and A. W. Lo, *Optimal control of execution costs*, Journal of Financial Markets, vol. 1, no. 1, pp. 1–50, 1998.
5. Á. Cartea, S. Jaimungal, and J. Penalva, *Algorithmic and High-Frequency Trading*, Cambridge University Press, 2015.
6. Á. Cartea, S. Jaimungal, and L. Sánchez-Betancourt, *Reinforcement learning for algorithmic trading*, in Machine Learning and Data Sciences for Financial Markets, Cambridge University Press, 2023.
7. J. Gatheral, *No-dynamic-arbitrage and market impact*, Quantitative Finance, vol. 10, no. 7, pp. 749–759, 2010.
8. I. Gibbs, and E. J. Candès, *Conformal inference for online prediction with arbitrary distribution shifts*, Journal of Machine Learning Research, vol. 25, no. 162, pp. 1–36, 2024.
9. B. Hambly, R. Xu, and H. Yang, *Recent advances in reinforcement learning in finance*, Mathematical Finance, vol. 33, no. 3, pp. 437–503, 2023.
10. J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, *Distribution-free predictive inference for regression*, Journal of the American Statistical Association, vol. 113, no. 523, pp. 1094–1111, 2018.
11. S. Lin, and P. A. Beling, *An end-to-end optimal trade execution framework based on proximal policy optimization*, in Proc. 29th International Joint Conference on Artificial Intelligence (IJCAI), pp. 4548–4554, 2020.
12. A. Macrì, and F. Lillo, *Reinforcement learning for optimal execution when liquidity is time-varying*, Applied Mathematical Finance, vol. 31, no. 5, pp. 312–342, 2024.
13. A. Micheli, and M. Monod, *Deep reinforcement learning for online optimal execution strategies*, arXiv:2410.13493, 2024.
14. P. Nagy, J. Calliess, and S. Zohren, *Asynchronous deep double dueling Q-learning for trading-signal execution in limit order book markets*, Frontiers in Artificial Intelligence, vol. 6, 1151003, 2023.
15. Y. Nevmyvaka, Y. Feng, and M. Kearns, *Reinforcement learning for optimized trade execution*, in Proc. 23rd International Conference on Machine Learning (ICML), pp. 673–680, 2006.
16. B. Ning, F. H. T. Lin, and S. Jaimungal, *Double deep Q-learning for optimal execution*, Applied Mathematical Finance, vol. 28, no. 4, pp. 361–380, 2021.
17. A. A. Obizhaeva, and J. Wang, *Optimal trading strategy and supply/demand dynamics*, Journal of Financial Markets, vol. 16, no. 1, pp. 1–32, 2013.
18. A. F. Perold, *The implementation shortfall: Paper versus reality*, Journal of Portfolio Management, vol. 14, no. 3, pp. 4–9, 1988.
19. Y. Romano, E. Patterson, and E. Candès, *Conformalized quantile regression*, Advances in Neural Information Processing Systems, vol. 32, pp. 3543–3553, 2019.
20. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, *Proximal policy optimization algorithms*, arXiv:1707.06347, 2017.
21. K. Stankevičiūtė, A. M. Alaa, and M. van der Schaar, *Conformal time-series forecasting*, Advances in Neural Information Processing Systems, vol. 34, pp. 6216–6228, 2021.
22. S. Sun, R. Wang, and B. An, *Reinforcement learning for quantitative trading*, ACM Transactions on Intelligent Systems and Technology, vol. 14, no. 3, pp. 1–29, 2023.
23. R. S. Sutton, and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed., MIT Press, 2018.
24. R. J. Tibshirani, R. Foygel Barber, E. Candès, and A. Ramdas, *Conformal prediction under covariate shift*, Advances in Neural Information Processing Systems, vol. 32, pp. 2530–2540, 2019.
25. V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*, Springer, 2005.
26. C. Xu, and Y. Xie, *Sequential predictive conformal inference for time series*, in Proc. 40th International Conference on Machine Learning (ICML), pp. 38707–38727, 2023.
27. Z. Yang, E. J. Candès, and L. Lei, *Bellman conformal inference: Calibrating prediction intervals for time series*, arXiv:2402.05203, 2024.
28. M. Zaffran, O. Féron, Y. Goude, J. Josse, and A. Dieuleveut, *Adaptive conformal predictions for time series*, in Proc. 39th International Conference on Machine Learning (ICML), pp. 25834–25866, 2022.
29. X. Zhou, W. Chen, and M. Xu, *Two kinds of learning algorithms for continuous-time VWAP targeting execution*, arXiv:2411.06645, 2024.