ONLINE FACTOR CORRECTION WITH DELAYED CONFORMAL FEEDBACK

Anonymous authors
Paper under double-blind review

ABSTRACT

Fixed financial factors can lose predictive value as market conditions change, motivating online adjustment of their influence on stock selection. Prediction intervals offer information for this adjustment, but coverage alone specifies neither an allocation rule nor its effect on investment performance. Delayed returns further complicate the link between calibration and decisions. We propose a factor-correction layer that connects these steps through an executable prediction target: each factor's standalone strategy return after transaction costs. A forecast of this target rewards expected payoff, while an online prediction radius discounts uncertainty before factor scores are combined. Three interleaved feedback sequences accommodate the return delay; a signed universal-portfolio update yields explicit coverage bounds that retain the calibration state inherited from warmup. The contribution is a specified connection between delayed calibration and factor allocation, evaluated with matched controls that distinguish forecasting gains from interval effects. On 157 public price and volume factors in Chinese and US equity universes, the correction retains positive net-return estimates relative to fixed coefficients and reduces the mean loss on the worst 5% of days relative to forecast-only allocation in both markets. Return intervals include zero, so these findings support an observed return–risk tradeoff rather than a confirmed return gain. The framework makes the financial role and statistical scope of online uncertainty explicit.

1 INTRODUCTION

A financial factor is a numerical feature used to rank stocks, such as recent price momentum or trading volume. A fixed library of such features provides several views of each stock, but their predictive value and trading costs change over time. Adjusting their coefficients can respond to those changes without replacing the underlying features or stock-selection rule. The question is how online uncertainty estimates should enter this adjustment.

Conformal prediction calibrates prediction sets by comparing forecasts with observed outcomes. Online conformal prediction updates these sets sequentially to track the fraction of outcomes they contain under changing data (Gibbs & Candès, 2021; Liu et al., 2026). This addresses a prediction objective. To use the resulting intervals for factor allocation, one must also specify which payoff is predicted, when it becomes observable, and how the interval changes a stock decision. These links matter in a trading system: costs depend on holdings, outcomes arrive after decisions, and the return of a combined stock ranking need not equal an average of the factors' standalone returns. Comparing only with fixed coefficients also leaves a central question unanswered: did interval information help, or did the underlying return forecast account for the improvement?

We propose an online correction at the factor-combination stage. Each factor first runs a standalone stock strategy under a common execution rule. Its net return, including transaction costs, is the prediction target. Using only released outcomes, we forecast that return and calibrate a prediction interval. The forecast rewards larger expected payoffs; the interval radius penalizes prediction uncertainty. Ranking these return–uncertainty scores produces positive, bounded relative coefficients for combining the original factor scores. Stock selection and execution then follow the common rule. This construction gives uncertainty a specified financial target and a specified route into decisions.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The calibration update builds on the universal-portfolio (UP) approach of Liu et al. (2026). It mixes bets on two artificial assets representing prediction misses and hits; their accumulated wealth determines the next interval radius. These calibration assets are distinct from the traded stocks. We retain a signed radius, with a negative value denoting an empty prediction set, so the coverage event and wealth recurrence agree even at zero prediction error. For delayed returns, three interleaved sequences, called feedback lanes, each process an outcome before issuing their next prediction. We derive explicit empirical coverage bounds both from initialization and for evaluation windows that inherit a calibration state from warmup. The contribution lies in this delayed, execution-based factor correction and its analysis; the underlying universal-portfolio calibration principle comes from prior work.

To identify the role of uncertainty, we compare three allocations with identical features and execution. *Fixed* uses unit factor coefficients; *Mean* adapts coefficients using the return forecast alone; *UP* adds the universal-portfolio interval penalty. Mean minus Fixed measures the forecast contribution, while UP minus Mean measures the incremental interval effect. This comparison is necessary because a coverage guarantee for factor returns does not imply a guarantee for the executed portfolio.

Across 157 public price and volume factors in Chinese and US equity universes, forecasting increases observed net return, and the interval penalty reduces worst-tail mean loss while giving up part of that increase. The full correction retains positive return estimates against Fixed in both markets, although the paired return intervals include zero. The study thus identifies a concrete return–risk tradeoff and the component responsible for each change. An exploratory extension weights calibration errors by each factor's influence on the current stock score; its analysis states what can be inferred when some factor outcomes are missing. The main contribution remains the connection between a delayed prediction target, an explicit allocation rule, and an evaluation that separates forecasting from uncertainty.

2 Related Work

Sequential uncertainty. Conformal prediction constructs prediction sets from a calibrated score (Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023). Conformalized quantile regression separates the predictive model from its calibration correction (Romano et al., 2019). Extensions treat covariate shift and nonexchangeable samples through additional assumptions and weighted scores (Tibshirani et al., 2019; Barber et al., 2023; Farinhas et al., 2024). Adaptive conformal inference, proportional–integral–derivative (PID) feedback, and decaying step sizes instead use sequential errors to track coverage (Gibbs & Candès, 2021; 2024; Angelopoulos et al., 2023; 2024b). Coin betting and universal portfolios reduce sensitivity to learning-rate choices (Podkopaev et al., 2024; Liu et al., 2026). Portfolio-based online conformal prediction (OCP) already extends to soft feature groups, whose memberships may be fractional (Bharti et al., 2026). We instead define the feedback unit as the stock score's sensitivity to factor coefficients at prediction time, distributed across simultaneously predicted factor returns. The resulting weighted average of their miss indicators can lie between zero and one. The regret notion matters for coverage claims (Ramalingam et al., 2025); we use the exact wealth of the specified update.

Prediction and decisions. A small prediction error and a useful downstream decision are different objectives (Elmachtoub & Grigas, 2022; Wilder et al., 2019). Conformal decision theory and utility-directed calibration connect uncertainty to a specified decision loss (Lekeufack et al., 2024; Cortes-Gomez et al., 2025). Expected risk control and high-probability population-risk control also concern different statistical quantities (Angelopoulos et al., 2024a; Bates et al., 2021). Our prediction sets concern factor-strategy returns; net investment performance is measured by replaying the resulting stock strategy.

Selection and aggregation. Selective prediction and conformal selection require explicit treatment of the selected population (El-Yaniv & Wiener, 2010; Geifman & El-Yaniv, 2019; Jin & Candès, 2023; Jin & Ren, 2025; Bao et al., 2025). Multivalid prediction pursues calibration across specified subgroups (Bastani et al., 2022). Conformal Online Model Aggregation combines prediction sets under a different aggregation contract (Gasparin & Ramdas, 2026). We combine stock-factor scores after calibrating their standalone payoffs. Our changing coefficients are related to

tracking useful experts over time (Herbster & Warmuth, 1998), but the stock ranking and holding constraints prevent identifying the executed payoff with an average expert return.

Quantitative evaluation. Qlib, a quantitative-investment research platform, provides public price/volume features and a daily stock-selection benchmark that holds up to k stocks and replaces a limited number at each rebalance (Yang et al., 2020). We use those public ingredients to isolate coefficient adaptation, rather than changing the return predictor and execution system together. Multiple factor and strategy comparisons can induce selection bias (White, 2000; Harvey et al., 2016; Bailey et al., 2017). We preserve all predefined methods and calendar periods in the appendix, use paired time-block summaries (Künsch, 1989), and distinguish descriptive intervals from multiplicity-adjusted diagnostics (Holm, 1979).

3 Online Allocation of Fixed Factors

Let t index market sessions, i stocks, and f factors. The correction changes factor coefficients before stocks are ranked. Its five stages are

$$z_{tif} \longmapsto (\mu_{tf}, q_{tf}) \longmapsto a_{tf} \longmapsto s_{ti}(a_t) \longmapsto (h_{t+1}, c_{t+1}), \tag{1}$$

Here z_{tif} is an oriented percentile factor score, μ_{tf} forecasts the factor strategy's net return, q_{tf} is its signed prediction radius, $a_{tf} > 0$ is its coefficient, and s_{ti} is the combined stock score. The vector a_t collects factor coefficients; h_{t+1} and c_{t+1} are post-trade stock holdings and cash at the next close. The feature definitions, score-combination rule, and execution operator remain fixed across interval methods.

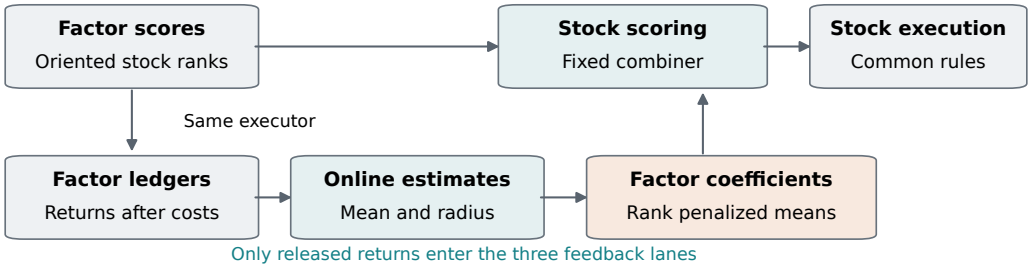

Figure 1: Factor scores feed the stock combiner and separate standalone strategy ledgers. Released net returns update forecasts and prediction radii, which set the coefficients supplied to the combiner. Fixed, Mean, and UP share features and execution; their coefficient inputs differ.

1. Fixed factor predictions. We start with the 157 price and volume features obtained by removing `VWAP0`, the relative volume-weighted average price feature, from Qlib's Alpha158 specification (a library of 158 daily price and volume features). Each feature uses adjusted daily observations through the current market session t. An orientation is estimated from its mean daily cross-sectional rank correlation with the stock return from close $t+1$ to close $t+2$, using 2008–2012 training examples whose outcome horizon ends within training. A negative mean reverses the feature; a zero mean keeps its sign. We retain features with at least 252 supported training dates, each using at least 30 stocks. Let $F \leq 157$ be the resulting fixed inventory for a market. The oriented feature is converted to an average percentile rank z_{tif} among available stocks in the contemporaneous universe. Stocks require at least $\lceil 0.8F \rceil$ available features. Missing feature values remain missing. Appendix G gives the feature and availability conventions.

2. Factor-level return and risk. Each factor runs a standalone strategy through the same stock-selection and execution operator used by the combined signal. Its independent holdings ledger starts in 2014. A signal issued after close t executes at close $t+1$ and is marked at close $t+2$, before the next rebalance. Write $B^{-}_{f,s}$ for its net asset value (NAV) immediately before trading at close s.

Define its target return Y_{tf} and absolute forecast-error score S_{tf} by

$$Y_{tf} = \frac{B^-_{f,t+2}}{B^-_{f,t+1}} - 1, \qquad S_{tf} = |Y_{tf} - \mu_{tf}|. \tag{2}$$

This return includes the fees paid at $t+1$, the strategy's existing positions, and their subsequent price changes. It is independent of the adaptive coefficients. The forecast μ_{tf} is stored at issue, before Y_{tf} is available. Returns and radii use decimal units; one basis point is 10^{-4}.

At session $t+3$, release the label before issuing the new forecast. Its mark date is strictly earlier than the new issue date. A finite label with complete portfolio marks updates a common factor mean,

$$\mu_f \leftarrow (1-\gamma)\mu_f + \gamma Y_{tf}, \qquad \gamma = 1 - 2^{-1/20}, \qquad \mu_f^{(0)} = 0. \tag{3}$$

The half-life is 20 observed returns. Incomplete labels update neither the mean nor the interval model. They remain visible in portfolio accounting. The same observed-label population is used for every interval method.

To preserve sequential feedback, assign each issue ordinal to one of three lanes by $t \bmod 3$. Each factor has three independent radius states, while its mean uses all released labels. The preceding label in a lane is processed, or resolved as missing, before that lane issues again. Thus each observed lane follows issue, observe, update, issue. Appendix C formalizes the event order and the treatment of missing and terminal labels.

Within one factor and lane, suppress t, f: let n index observed feedback, N_{n-1} count prior misses, and W_{n-1} be prior artificial betting wealth. Write α for the target miss rate, λ_n for the predictive betting fraction, and m_n for the miss indicator, with $\mathbf{1}\{\cdot\}$ equal to one when its event holds. The signed UP construction gives (Liu et al., 2026):

$$\lambda_n = \frac{N_{n-1} + 1/2}{n}, \qquad q_n = W_{n-1}\frac{\lambda_n - \alpha}{\alpha(1-\alpha)}, \tag{4}$$

$$m_n = \mathbf{1}\{S_n > q_n\}, \qquad W_n = W_{n-1} - (\alpha - m_n)q_n. \tag{5}$$

After feedback, set $N_n = N_{n-1} + m_n$. The fixed parameters are $\alpha = 0.1$, initial wealth $W_0 = \epsilon = 0.01$, and $N_0 = 0$. The issued set is $[\mu_{tf} - q_{tf}, \mu_{tf} + q_{tf}]$ when $q_{tf} \geq 0$ and the empty set when $q_{tf} < 0$. Coverage and the wealth update use this same signed threshold, including at zero residual. The two assets in the UP construction encode calibration errors; they are distinct from the financial factors. Appendix B derives the update and its relationship to the versioned UP-OCP algorithm.

3. Dynamic factor coefficients. Define the UP allocation score u_{tf} and its positive coefficient by

$$u_{tf} = \mu_{tf} - \tfrac{1}{4}\max(q_{tf}, 0), \qquad a_{tf} = \exp\{2(r_{tf} - 1/2)\}, \tag{6}$$

where r_{tf} is the average percentile rank of u_{tf} among factors with at least 60 released complete labels. An unready factor receives coefficient one. All coefficients equal one during the 2014 warmup; adaptation starts in 2015. Coefficients are positive and their largest ratio is below e^2. The multiplier changes relative influence rather than removing a factor. The penalty $1/4$ is a fixed risk preference. It does not turn u into a confidence bound on a conditional mean. A negative signed radius is projected to zero only for this allocation score; the calibration update remains signed.

The radius measures absolute forecast error, not the portfolio's return variability, decline from peak wealth, or mean loss on its worst days. Penalizing it favors factor payoffs that are easier to forecast, which can reduce downstream risk after aggregation, but that reduction is not supplied by the coverage theorem. Financial risk is therefore measured on the executed portfolio and attributed against both Fixed and Mean.

The controls separate the two components of the correction. Fixed sets every coefficient to one. Mean ranks μ alone. Rolling and the Krichevsky–Trofimov (KT) control replace the UP radius in Equation 6: Rolling uses the upper empirical 90th percentile of the last 60 released absolute errors, and KT uses the symmetric bettor of Podkopaev et al. (2024). The UP-ratio control ranks $\mu_{tf}/\max(q_{tf}, 10^{-4})$. They share the mean forecast, readiness rule, and coefficient map. UP versus Fixed measures the complete correction; UP versus Mean isolates the addition of interval information.

4. Fixed score combination. Let O_{ti} be the set of factors available for stock i at session t. The primary combiner, called All, uses the stock score

$$s_{ti}(a_t) = \frac{\sum_{f \in O_{ti}} a_{tf} z_{tif}}{\sum_{f \in O_{ti}} a_{tf}}. \tag{7}$$

Both the numerator and denominator use that stock's actual feature support. The availability mask is common across methods. Unit coefficients recover the fixed equal-feature score, and multiplying every coefficient by the same positive constant leaves stock scores unchanged. A second combiner, Family, first normalizes coefficients within each of six fixed feature groups, then averages the eligible group scores. Appendix D.2 defines its groups and availability rule; both combiners use the same executor.

5. Fixed execution. The combined score enters Top50/Drop5: a deterministic rule that holds up to 50 stocks and proposes up to five replacements at each rebalance. It retains positions between rebalances, uses a 95% investment budget, and breaks ties by a stable stock identifier. Each method has its own portfolio state and the same execution rules. Orders use the next session's close, tradability filter, share-unit conversion, and transaction-cost model. Rejected orders generate neither a fill nor fictional cash. The base costs are 5 basis points on buys and 15 on sells, with a minimum of five currency units per executed order. Holdings are marked through the last supported close; there is no artificial terminal sale. Appendix A specifies the common NAV and cost identities. The method allocates factor influence before stock selection; it does not trade a linear mixture of standalone portfolios.

3.1 What the Calibration Bound Controls

The following specialization uses the standard Jeffreys mixture identity and binary Pinsker inequality. Appendix B gives a complete proof for the signed recurrence.

Proposition 1 (Coverage on a scored prefix)**.** *For any finite nonnegative scores $S_1, \dots, S_T$ in one lane, $T \geq 1$, let $p_T = T^{-1} \sum_{n=1}^{T} m_n$. Equations 4–5 give positive wealth and*

$$|p_T - \alpha| \leq \sqrt{\frac{1}{2T}\left[\log\left(1 + \frac{(1-\alpha)\sum_{n=1}^{T} S_n}{\epsilon}\right) + \log\frac{\pi(T+1)}{2}\right]}. \tag{8}$$

The bound is pathwise: it holds for each realized score sequence. It bounds the empirical miss rate since that lane's initialization. Under polynomial score growth it tends to zero. Pooling lanes uses their observed counts and requires no independence between lanes. For a test window entered with a warmup state, the prior and wealth at that boundary replace the fresh initialization; the exact window bound appears in Appendix C.2. These statements concern empirical prediction-set coverage. Executed net return, drawdown, and tail loss are evaluated separately on the complete strategy trajectories.

3.2 Decision-Weighted Fractional Feedback

Factorwise coverage treats each factor separately, regardless of its influence on the stock ranking. An exploratory extension, influence-weighted universal-portfolio feedback (IUP), instead weights each factor's prediction miss by the stock score's sensitivity to that factor's coefficient. At issue, differentiate the All score with respect to each log coefficient, evaluated at Mean's coefficients. Average the absolute derivatives over eligible stocks and normalize across factors to obtain $v_{tf} \geq 0$, $\sum_f v_{tf} = 1$. Unavailable factors contribute zero derivative; if all derivatives vanish, use $v_{tf} = 1/F$. These weights measure score sensitivity, not marginal contributions to portfolio return.

Let σ_{tf} be the exponentially weighted moving average (EWMA) of released absolute errors, with half-life 20, initial value .01, and floor 10^{-4}. Standardized errors are $X_{tf} = |Y_{tf} - \mu_{tf}|/\sigma_{tf}$. For the set O_t of complete released factor labels, define observed weight $H_t = \sum_{f \in O_t} v_{tf}$ and $w_{tf} = v_{tf}/H_t$. If $H_t > 0$, the fractional miss against the common signed radius q_t is

$$m_t = \sum_{f \in O_t} w_{tf} \mathbf{1}\{X_{tf} > q_t\} \in [0, 1]. \tag{9}$$

Issues with $H_t = 0$ do not update calibration. Each of the three lanes mixes constant two-asset bets using this fractional feedback. The allocation score is $\mu_{tf} - .25\sigma_{tf} \max(q_t, 0)$, with the original ranking map and executor. Uniform-weighted universal-portfolio feedback (UUP) uses $v_{tf} = 1/F$ as an ablation.

Appendix F gives the recurrence and proves a finite-sample certificate for the observed weighted miss sequence. Missing outcomes limit its scope: for $H_t > 0$, the full-factor miss lies within $[H_t m_t, H_t m_t + 1 - H_t]$. The certificate does not bound portfolio tail risk. The underlying portfolio calibration and soft-group ideas are established (Liu et al., 2026; Bharti et al., 2026); this extension specifies a decision-based feedback distribution over simultaneous factor returns and makes its observation boundary explicit.

4 Public-Market Evaluation

4.1 Data, Comparators and Chronology

We evaluate the same system on two public Qlib-format archives: Chinese A-shares with historical China Securities Index 300 (CSI300) membership and US stocks with historical Standard & Poor's 500 (S&P 500) membership. The files are distributed through the SunsetWolf Qlib dataset releases used by Qlib's public downloader. Qlib documents the underlying daily price and volume collection as Yahoo Finance data. The archive supplies adjusted open, high, low, close, volume and adjustment factors, but no volume-weighted average price (VWAP). We therefore use the same 157-feature subset in both markets. Appendix G records the supplier chain, snapshot identifiers, historical universe files, feature definitions and data-quality audit.

The chronology separates feature orientation from adaptive allocation. We use 2007 for feature history, 2008–2012 for factor signs, and reserve 2013 for development. No parameter is selected from 2013 strategy returns. In 2014, fixed coefficients establish portfolio holdings and calibration histories. The test runs from the first eligible 2015 signal through the final signal whose return is marked by 31 July 2020: 1,325 daily returns in China and 1,403 in the US. Calendar-year summaries use this continuing trajectory; 2020 is a seven-month block. The six calendar blocks cover more than five years of trading in each market and include both quiet periods and large market declines.

Every comparison fixes the oriented features, available-stock mask, signal normalization, Top50/Drop5 operator and transaction costs. Fixed is the direct baseline: all factor coefficients equal one before the stock ranking. Mean adds dynamic weights based only on the causal factor-return forecast. The four interval allocations add Rolling, KT, UP or UP-ratio uncertainty scores. The main comparisons use All. Its emphasis was chosen after observing both combiners; complete Family results remain in Appendix H. This design measures the complete correction against Fixed and its uncertainty component against Mean. It does not change the stock-selection procedure between these comparators.

We report net daily means in basis points (bp), compound annual growth rates (CAGR) using 252 sessions, annualized mean-to-standard-deviation ratios with zero risk-free return (Sharpe), maximum peak-to-trough drawdown (MDD), and expected shortfall ($ES_{.05}$), the mean loss on the worst 5% of days. Returns include executed fees. A second replay doubles the fees while retaining the base-cost forecast and coefficient sequences, isolating execution-cost sensitivity. Cash earns zero. The simulation uses daily closes and does not model market impact or additional slippage.

For each paired difference, we resample the daily difference series with circular blocks of 20 sessions, using 2,000 resamples and a fixed seed. The reported 95% percentile intervals describe sampling variability along the observed path. Holm-adjusted bootstrap sign diagnostics compare the adaptive methods within a market and cost setting. All method and annual results appear in the appendix; no test-period coefficient, threshold or hyperparameter is optimized.

For interpretation, let $\bar{R}_0$, $\bar{R}_\mu$ and $\bar{R}_{\mathrm{UP}}$ denote the realized net means of Fixed, Mean and UP. The return attribution is the exact identity

$$\underbrace{\bar{R}_{\mathrm{UP}} - \bar{R}_0}_{\Delta_{\mathrm{total}}} = \underbrace{\bar{R}_\mu - \bar{R}_0}_{\Delta_{\mathrm{forecast}}} + \underbrace{\bar{R}_{\mathrm{UP}} - \bar{R}_\mu}_{\Delta_{\mathrm{interval}}}. \tag{10}$$

Table 1: Public-market results, January 2015–July 2020, base transaction costs. The baseline uses fixed equal-feature coefficients. CAGR and downside losses are percentages; net means are basis points per day.

Market	Method	Net bp/d	CAGR %	Sharpe	MDD %	$ES_{.05}$ %
China	Fixed	2.16	1.91	0.20	61.08	4.41
China	Mean	3.66	5.71	0.34	56.85	4.45
China	Rolling	3.19	4.57	0.30	56.52	4.41
China	KT	2.83	3.70	0.27	59.65	4.35
China	UP	3.35	5.14	0.32	53.21	4.32
China	UP-ratio	3.36	4.87	0.31	58.22	4.50
US	Fixed	2.12	2.36	0.22	51.95	3.56
US	Mean	2.93	4.38	0.30	51.97	3.57
US	Rolling	2.56	3.95	0.29	49.81	3.27
US	KT	2.90	4.73	0.32	49.19	3.34
US	UP	2.48	3.46	0.26	51.87	3.44
US	UP-ratio	2.67	3.72	0.27	52.60	3.60

For a lower-is-better risk functional D, the interval-induced risk shift is $D_{\mathrm{UP}} - D_{\mu}$. These quantities separate the complete correction from the marginal effect of interval information. They are an attribution of the fixed replay, not a replacement for its predeclared success criterion.

4.2 Net Return and Downside Risk

The correction exhibits the same two-stage pattern in both markets (Table 1). Forecasting supplies the return gain: Mean minus Fixed is $+1.50$ bp/day in China and $+0.81$ bp/day in the US. Charging the forecast for the UP radius gives back 0.31 and 0.45 bp/day, respectively. The complete UP correction nevertheless remains above Fixed by 1.19 and 0.36 bp/day. Chinese CAGR rises from 1.91% under Fixed to 5.14% under UP; US CAGR rises from 2.36% to 3.46%.

The interval charge changes the downside profile. Relative to Mean, UP lowers $ES_{.05}$ by 0.126 percentage points in China and 0.123 points in the US, and lowers annualized volatility from 26.94% to 26.11% and from 24.79% to 23.76%. Chinese maximum drawdown falls from 56.85% to 53.21%; US drawdown is essentially unchanged at 51.97% versus 51.87%. Relative to Fixed, UP keeps the positive return point estimate while lowering ES in both markets and maximum drawdown by 7.87 points in China. The empirical correction therefore uses part of the forecast gain to improve selected downside measures rather than adding further mean return.

An additional paired-block diagnostic quantifies the risk differences (Appendix H.17). UP minus Fixed drawdown intervals include zero in both markets: $[-10.44, 1.44]$ percentage points in China and $[-5.35, 0.63]$ in the US. The US expected-shortfall difference is -0.122 percentage points with interval $[-0.204, -0.052]$. This exploratory risk analysis was added after the predefined financial replay. Against Mean, the ES differences are -0.126 points with interval $[-0.224, -0.030]$ in China and -0.123 with $[-0.216, -0.041]$ in the US. These unadjusted diagnostics support the tail-risk side of the attribution but do not change the predeclared joint efficacy criterion.

Table 2 quantifies the paired comparisons. UP minus Fixed is 1.19 bp/day in China, with interval $[-0.27, 2.72]$, and 0.36 bp/day in the US, with interval $[-0.48, 1.11]$. Neither lower endpoint is positive. Doubling fees preserves the signs of the pooled point estimates and leaves both intervals crossing zero. In China, UP's lower turnover reduces mean daily fees from 1.23 to 1.10 bp relative to Fixed, explaining part of the stronger comparison at doubled costs. The common-coefficient cost replay does not refit the estimator on high-cost labels.

The annual decomposition makes the time variation explicit (Figure 6). UP exceeds Fixed in five of the six calendar blocks in each market. Its largest paired gains occur in 2018: 4.85 bp/day in China and 1.49 bp/day in the US. Every annual UP–Fixed interval includes zero. Against Mean, the 2020 Chinese block loses 4.80 bp/day, with interval $[-7.92, -1.58]$; the corresponding US difference is -3.41 bp/day with interval $[-7.74, 0.45]$. The pooled gain over Fixed therefore combines substantial annual variation with an unresolved incremental return benefit from the interval.

Table 2: Paired UP net-return gains against Fixed and Mean. All entries use the same period as Table 1; intervals use time-block resampling. Cost multipliers apply to both proportional and minimum order fees.

Market	Costs	Comparator	Δ bp/d	95% block interval
China	1×	Fixed	+1.19	[-0.27, 2.72]
China	1×	Mean	-0.31	[-1.67, 1.09]
US	1×	Fixed	+0.36	[-0.48, 1.11]
US	1×	Mean	-0.45	[-1.35, 0.39]
China	2×	Fixed	+1.32	[-0.17, 2.89]
China	2×	Mean	-0.17	[-1.55, 1.25]
US	2×	Fixed	+0.36	[-0.48, 1.12]
US	2×	Mean	-0.45	[-1.35, 0.39]

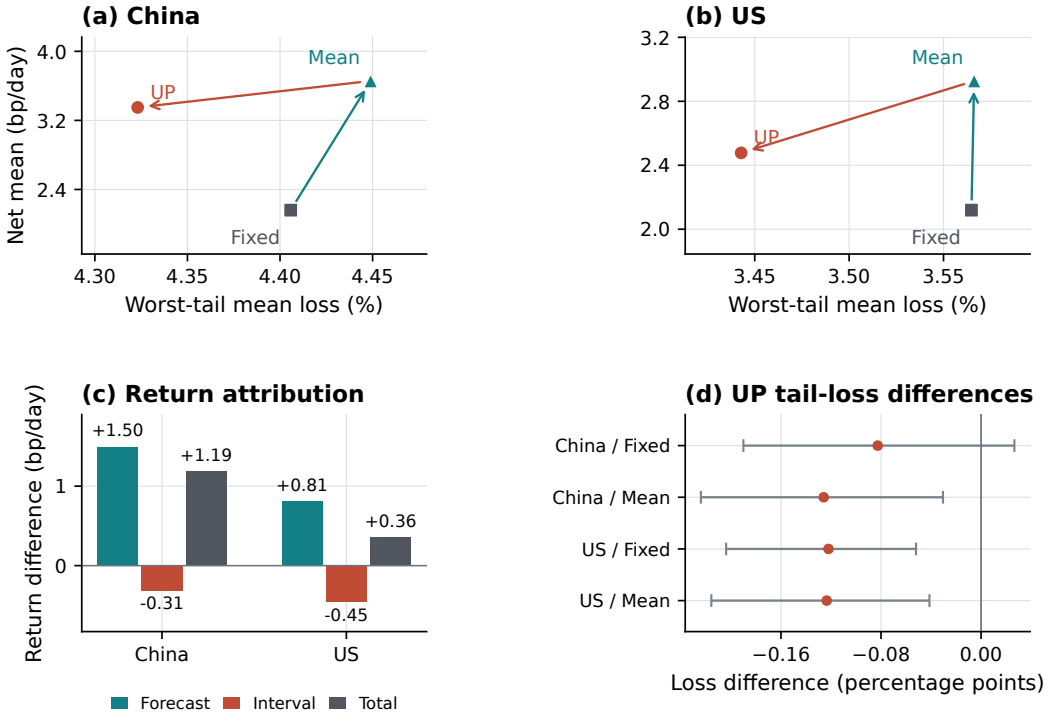

Figure 2: Return and downside-risk attribution at base costs. Top: realized net mean against worst-5% mean loss; arrows show Fixed → Mean and Mean → UP. Bottom left: the exact return decomposition in Equation 10; the bars show the forecast gain, interval effect, and total correction. Bottom right: exploratory paired UP differences in expected shortfall with 95% block intervals; negative values denote lower tail loss.

4.3 Coverage, Observation and Execution

The observed UP coverage is 83.54% in China and 86.23% in the US (Table 3). Rolling is closer to 90% in both markets, at 88.81% and 89.35%. The signed UP and KT games start separately in three lanes, and each lane receives only its complete labels. The finite-window guarantee in Appendix C.2 accounts for these sample sizes and the wealth inherited from warmup. It does not require a finite test window to achieve exactly 90% empirical coverage.

Missing observations are part of the experiment. A factor label is excluded from estimation whenever an existing position lacks an entry or terminal mark. Only 54,809 of 208,025 Chinese factor labels and 143,918 of 220,271 US labels are complete; the ready subset contains 53,230 and 143,895 labels. Calibration is measured on this observed subset. Portfolio returns continue to include every test day, carrying the last observed mark when necessary. The appendix reports the affected portfolio

Table 3: Calibration on complete factor labels after the 60-label readiness threshold. Coverage targets 90%; widths refer to the realized prediction sets, in basis points of the standalone strategy return.

Market	Update	Coverage %	Mean width bp	Observed labels
China	Rolling	88.81	451.5	53,230
China	KT	81.75	354.8	53,230
China	UP	83.54	440.8	53,230
US	Rolling	89.35	330.5	143,895
US	KT	85.35	267.8	143,895
US	UP	86.23	313.4	143,895

weight, observed-price fractions and complete annual calibration results. This distinction prevents a high observed-label coverage from being interpreted as coverage of unavailable outcomes.

We verify the signal-to-label delay by perturbing future returns and checking that earlier coefficients remain identical. All dynamic methods exactly recover their Fixed portfolio during warmup. An independent reconciliation checks continuing NAV, fees, reported metrics and the native UP wealth sequence. A price-unit regression test also verifies that rescaling adjusted prices and share units preserves orders and returns. These checks establish the executed comparison; they do not certify the upstream archive's corporate-action history.

4.4 Exploratory Decision-Weighted Extension

We froze IUP and its uniform-feedback ablation UUP before their outcomes were examined, then replayed both markets, both score combiners and two cost settings. The 2015–2020 test outcomes of the older methods were already known; these are retrospective mechanism tests, not independent confirmation. All six older Fixed, Mean and UP portfolios per cost setting match their stored ledgers exactly. Appendix I preserves every annual comparison and the risk intervals. On the same released issues and with the same score-sensitivity weights, the observed miss rate falls from 17.75% under factorwise UP to 13.95% under IUP in China, and from 13.66% to 12.94% in the US. Neither reaches the 10% target. The finite-prefix certificate holds in each lane and is loose: its absolute-error allowance is about 12.2 percentage points in China and 11.4 in the US. Under arbitrary missing-factor outcomes, the reported outer bounds on the full-factor miss rate are $[4.08\%, 74.67\%]$ in China and $[8.26\%, 44.06\%]$ in the US. There are also 144 Chinese test issues without scored feedback and three terminal US issues.

Pooled over 2,728 daily decisions with separate market resampling, IUP–Fixed under All is $+0.58$ bp/day $[-0.38, 1.63]$; IUP–Mean is -0.56 $[-1.51, 0.39]$. Under Family, IUP–Fixed is -0.76 $[-1.92, 0.28]$. The US All expected-shortfall shift versus Mean is -0.280 percentage points $[-0.494, -0.098]$, whereas the Chinese shift is effectively zero $[-0.215, 0.217]$. These intervals are exploratory and unadjusted. The extension improves an observed coverage diagnostic but neither demonstrates additional net return over Mean nor establishes a cross-market downside-risk reduction.

5 Discussion and Conclusion

The original replay orders net return as Mean $>$ UP $>$ Fixed in both markets. Mean supplies the point-estimate gain; the UP interval charge spends part of it and lowers exploratory expected shortfall against Mean. Yet every annual UP–Fixed return interval crosses zero. The predeclared joint gate fails because UP does not beat Mean in return and its paired lower bound against Fixed is not positive. The evidence supports an observed return–risk tradeoff; it does not establish an incremental return benefit from interval information.

IUP improves observed calibration but does not beat Mean in pooled net return. Missing marks leave full-factor coverage weakly identified, and neither wealth certificate bounds portfolio tail risk. Source and execution audits, including every negative comparison, remain in the appendix.

AI USE STATEMENT

Generative AI assisted literature search, mathematical and experimental development, implementation and synthetic verification, data processing and analysis, manuscript drafting and editing, and figure preparation.

REPRODUCIBILITY STATEMENT

The accompanying code, fixed protocol, data acquisition instructions and complete result tables reproduce the public-data study. The appendix specifies the feature definitions, chronological release rules, execution ledger and all comparisons. Raw market observations are obtained from the cited upstream archives.

REFERENCES

Anastasios N. Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023. doi: 10.1561/2200000101.

Anastasios N. Angelopoulos, Emmanuel J. Candès, and Ryan J. Tibshirani. Conformal PID control for time series prediction. In *Advances in Neural Information Processing Systems*, volume 36, pp. 23047–23074, 2023. doi: 10.52202/075280-1000. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/47f2fad8c1111d07f83c91be7870f8db-Abstract.html.

Anastasios N. Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. In *The Twelfth International Conference on Learning Representations*, 2024a. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/f3549ef9b5ff520a7e41ff3cc306ab2b-Abstract-Conference.html.

Anastasios Nikolas Angelopoulos, Rina Foygel Barber, and Stephen Bates. Online conformal prediction with decaying step sizes. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 1616–1630, 2024b. URL https://proceedings.mlr.press/v235/angelopoulos24a.html.

David H. Bailey, Jonathan M. Borwein, Marcos López de Prado, and Qiji Jim Zhu. The probability of backtest overfitting. *The Journal of Computational Finance*, 20(4), 2017. doi: 10.21314/JCF.2016.322. URL https://www.risk.net/journal-of-computational-finance/2471206/the-probability-of-backtest-overfitting.

Yajie Bao, Yuyang Huo, Haojie Ren, and Changliang Zou. CAP: A general algorithm for online selective conformal prediction with FCR control. *Journal of Machine Learning Research*, 26 (287):1–74, 2025. URL https://jmlr.org/papers/v26/24-0452.html.

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023. doi: 10.1214/23-AOS2276.

Osbert Bastani, Varun Gupta, Christopher Jung, Georgy Noarov, Ramya Ramalingam, and Aaron Roth. Practical adversarial multivalid conformal prediction. In *Advances in Neural Information Processing Systems*, volume 35, pp. 29362–29373, 2022. doi: 10.52202/068431-2129. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/bcdaaa1aec3ae2aa39542acefdec4e4b-Abstract.html.

Stephen Bates, Anastasios Angelopoulos, Lihua Lei, Jitendra Malik, and Michael I. Jordan. Distribution-free, risk-controlling prediction sets. *Journal of the ACM*, 68(6):1–34, 2021. doi: 10.1145/3478535. URL https://dl.acm.org/doi/10.1145/3478535.

Beepul Bharti, Ambar Pal, Jacopo Teneggi, and Jeremias Sulam. Parameter-free and group conditional online conformal prediction, 2026. URL https://arxiv.org/abs/2606.00419v4. Version 4, August 2026.

Santiago Cortes-Gomez, Carlos Patiño, Yewon Byun, Zhiwei Steven Wu, Eric Horvitz, and Bryan Wilder. Utility-directed conformal prediction: A decision-aware framework for actionable uncertainty quantification. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/0c6b452f1bbfb6905f6bac957d73b321-Abstract-Conference.html.

Ran El-Yaniv and Yair Wiener. On the foundations of noise-free selective classification. *Journal of Machine Learning Research*, 11(53):1605–1641, 2010. URL https://jmlr.org/papers/v11/el-yaniv10a.html.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1): 9–26, 2022. doi: 10.1287/mnsc.2020.3922. URL https://pubsonline.informs.org/doi/10.1287/mnsc.2020.3922.

António Farinhas, Chrysoula Zerva, Dennis Ulmer, and André F. T. Martins. Non-exchangeable conformal risk control. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/de04896f011beff76c91e094f72727f4-Abstract-Conference.html.

Matteo Gasparin and Aaditya Ramdas. Conformal online model aggregation. In *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 1663–1683, 2026. URL https://proceedings.mlr.press/v337/gasparin26a.html.

Yonatan Geifman and Ran El-Yaniv. SelectiveNet: A deep neural network with an integrated reject option. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2151–2159, 2019. URL https://proceedings.mlr.press/v97/geifman19a.html.

Isaac Gibbs and Emmanuel J. Candès. Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems*, volume 34, pp. 1660–1672, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/0d441de75945e5acbc865406fc9a2559-Abstract.html.

Isaac Gibbs and Emmanuel J. Candès. Conformal inference for online prediction with arbitrary distribution shifts. *Journal of Machine Learning Research*, 25(162):1–36, 2024. URL https://jmlr.org/papers/v25/22-1218.html.

Campbell R. Harvey, Yan Liu, and Heqing Zhu. ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68, 2016. doi: 10.1093/rfs/hhv059. URL https://academic.oup.com/rfs/article/29/1/5/1843824.

Mark Herbster and Manfred K. Warmuth. Tracking the best expert. *Machine Learning*, 32(2): 151–178, 1998. doi: 10.1023/A:1007424614876. URL https://link.springer.com/article/10.1023/A:1007424614876.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. URL https://www.jstor.org/stable/4615733.

Ying Jin and Emmanuel J. Candès. Selection by prediction with conformal p-values. *Journal of Machine Learning Research*, 24(244):1–41, 2023. URL https://jmlr.org/papers/v24/22-1176.html.

Ying Jin and Zhimei Ren. Confidence on the focal: conformal prediction with selection-conditional coverage. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(4):1239–1259, 2025. doi: 10.1093/jrsssb/qkaf016. URL https://academic.oup.com/jrsssb/article/87/4/1239/8113856.

Hans R. Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989. doi: 10.1214/aos/1176347265.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113 (523):1094–1111, 2018. doi: 10.1080/01621459.2017.1307116. URL https://www.stat.berkeley.edu/~ryantibs/papers/conformal.pdf.

Jordan Lekeufack, Anastasios N. Angelopoulos, Andrea Bajcsy, Michael I. Jordan, and Jitendra Malik. Conformal decision theory: Safe autonomous decisions from imperfect predictions. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11668–11675. IEEE, 2024. doi: 10.1109/ICRA57147.2024.10610041. URL https://ieeexplore.ieee.org/document/10610041.

Tuo Liu, Edgar Dobriban, and Francesco Orabona. Online conformal prediction via universal portfolio algorithms, 2026. URL https://arxiv.org/abs/2602.03168v1. Version 1.

Aleksandr Podkopaev, Dong Xu, and Kuang-Chih Lee. Adaptive conformal inference by betting. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 40886–40907. PMLR, 2024. URL https://proceedings.mlr.press/v235/podkopaev24a.html.

Ramya Ramalingam, Shayan Kiyani, and Aaron Roth. The relationship between no-regret learning and online conformal prediction. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 51060–51078, 2025. URL https://proceedings.mlr.press/v267/ramalingam25a.html.

Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/5103c3584b063c431bd1268e9b5e76fb-Abstract.html.

Ryan J. Tibshirani, Rina Foygel Barber, Emmanuel J. Candès, and Aaditya Ramdas. Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/8fb21ee7a2207526da55a679f0332de2-Abstract.html.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005. doi: 10.1007/b106715. URL https://link.springer.com/book/10.1007/b106715.

Halbert White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2000. doi: 10.1111/1468-0262.00152. URL https://onlinelibrary.wiley.com/doi/10.1111/1468-0262.00152.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1658–1665, 2019. doi: 10.1609/aaai.v33i01.33011658. URL https://ojs.aaai.org/index.php/AAAI/article/view/3982.

Xiao Yang, Weiqing Liu, Dong Zhou, Jiang Bian, and Tie-Yan Liu. Qlib: An AI-oriented quantitative investment platform, 2020. URL https://arxiv.org/abs/2009.11189. Preprint.

A Common NAV and Cost Accounting

Fix one market and one portfolio. Let h_s^-, c_s^- and h_s^+, c_s^+ denote holdings and cash immediately before and after trading at close s. Let $\widetilde{P}_s$ be the accounting mark, equal to the observed adjusted close when available and to the last observed mark for a carried position otherwise. Write D_s^+ and D_s^- for executed buy and sell notionals and K_s for their total fees. At the common close-price valuation,

$$B_s^- = c_s^- + (h_s^-)^\top \widetilde{P}_s, \tag{11}$$

$$c_s^+ = c_s^- + D_s^- - D_s^+ - K_s, \qquad B_s^+ = c_s^+ + (h_s^+)^\top \widetilde{P}_s = B_s^- - K_s. \tag{12}$$

Cash earns zero in this benchmark. Marking the post-trade holdings at the next close, before its trade, gives

$$\frac{B_{s+1}^-}{B_s^-} - 1 = \frac{(h_s^+)^\top(\widetilde{P}_{s+1} - \widetilde{P}_s)}{B_s^-} - \frac{K_s}{B_s^-}. \tag{13}$$

For decision t, set $s = t + 1$. Equation 13 proves Equation 2 and counts the entry/rebalance fee exactly once. Using B_s^+ as the denominator would remove that fee from the label unless it were accounted for separately. The trade at $s + 1$ belongs to the next label.

For an executed order of notional $v > 0$, the base fee is $\max(5, 0.0005v)$ on a buy and $\max(5, 0.0015v)$ on a sell. Zero executed quantity incurs zero fee. Notionals, the minimum fee, and initial capital 10^8 use the market's local currency. Actual buy-plus-sell notional divided by pre-trade NAV defines turnover. The same bookkeeping is used for every standalone factor and every combined method. The benchmark has no additional spread, slippage, capacity, or cash-interest term.

Share rounding uses the data's adjustment factor. If $P_{si}^{\mathrm{adj}} = b_{si} P_{si}^{\mathrm{raw}}$ with $b_{si} > 0$, then $h_{si}^{\mathrm{raw}} = b_{si} h_{si}^{\mathrm{adj}}$ preserves position value. An actual lot of L shares therefore corresponds to L/b_{si} adjusted shares. The lot is 100 for the China benchmark and one for the US benchmark. This conversion must be applied before checking order notional and fees. Prices, volume, and adjustment factors must be finite and positive for an execution. China additionally uses the fixed absolute daily-change filter $|P_s/P_{s-1} - 1| < 0.095$, where P_s is the observed adjusted close. This is the benchmark's common tradability convention.

The Top50/Drop5 operator retains positions unless selected for replacement. It uses the stored issue-time score and a deterministic stock-key tie-break. It considers high-ranked unheld candidates together with held stocks, selects up to five replacements, executes permitted sells, and caps new buys by actual remaining capacity. New purchases share the available investment budget, up to 95% of pre-trade NAV, after retained positions. Quantities are determined at the execution close. This is a close-price target-budget simulation; neither stock scores nor their ordering are recomputed with the execution-day features. Failed sells do not finance buys. Holdings remain nonnegative, cash is nonnegative to numerical tolerance, and no portfolio holds more than 50 stocks.

Accounting marks and observed outcomes are recorded separately. A carried unobserved mark at either endpoint makes the affected factor label incomplete; its value cannot enter the mean or radius update. The portfolio date and its unobserved weight remain in financial evaluation. At the terminal close, holdings are marked without an invented liquidation. Every method maintains its own cash and holdings from the same initial conditions. Consequently, changed coefficients can alter later turnover and costs even when the executor is fixed.

B Signed UP: Complete Wealth and Coverage Proof

The two-asset reduction and Jeffreys mixture are inherited from Liu et al. (2026), arXiv version 1. We specialize them to the signed recurrence in Equations 4–5. The mixture identity and binary Pinsker inequality below are standard facts, proved here to specify constants and endpoint conventions. The delayed-lane and continued-window statements are direct extensions for our observation schedule.

The displayed UP-OCP algorithm projects its radius to zero. Our reported set and update instead use the signed radius: a negative radius denotes an empty set. This distinction matters at a zero residual.

A negative signed radius misses zero, whereas a closed singleton centered on the prediction covers it. Keeping the event signed makes the set, gradient, and wealth recursion agree. The inspected reference code at commit `3575f1f94a14620e089422f52cf9e1a6766ec7b1` also changes the requested α by -0.015 and initializes radius zero; neither modification is used here. We derive the bound from wealth directly rather than transferring a constant from a different initialization or clipping convention.

B.1 Positive wealth and the exact mixture

For this subsection, suppress factor and lane indices. Assume $0 < \alpha < 1$, $W_0 = \epsilon > 0$, and finite scores $S_n \geq 0$. The index n advances only on an observed update. Before its outcome, $0 \leq N_{n-1} \leq n-1$, so $0 < \lambda_n < 1$. Substituting the definition of q_n into the wealth update gives

$$\frac{W_n}{W_{n-1}} = \begin{cases} \lambda_n/\alpha, & m_n = 1, \\ (1-\lambda_n)/(1-\alpha), & m_n = 0. \end{cases} \tag{14}$$

Both factors are strictly positive. This proves $W_n > 0$ at every finite step in exact arithmetic. With $N = \sum_{n=1}^{T} m_n$ and $H = T - N$, multiplying the successive miss and hit pseudo-counts gives

$$\prod_{n=1}^{T} \lambda_n^{m_n}(1-\lambda_n)^{1-m_n} = \frac{\Gamma(N+1/2)\Gamma(H+1/2)}{\pi\Gamma(T+1)} = \frac{\mathrm{B}(N+1/2, H+1/2)}{\pi}, \tag{15}$$

$$\frac{W_T}{\epsilon} = \frac{\mathrm{B}(N+1/2, H+1/2)}{\pi\alpha^N(1-\alpha)^H}. \tag{16}$$

Here Γ is the gamma function, and $\mathrm{B}(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ is the beta function. Although the likelihood identity depends only on the miss count, the radii and errors generating those misses depend on their temporal order. No assumption of independent misses has been made.

The wealth state has the same units as S and q. Equivalently one may store $U_n = W_n/\epsilon$, initialize $U_0 = 1$, and output $q_n = \epsilon U_{n-1}(\lambda_n - \alpha)/[\alpha(1-\alpha)]$. The two conventions are identical; an additional multiplication by ϵ when storing physical wealth would define a different update.

B.2 Finite-prefix discrepancy

On a hit, $q_n \geq S_n \geq 0$, and hence $W_n - W_{n-1} = -\alpha q_n \leq 0$. On a miss, $W_n - W_{n-1} = (1-\alpha)q_n \leq (1-\alpha)S_n$, including when $q_n < 0$. Summation yields the deterministic inequality

$$0 < W_T \leq \epsilon + (1-\alpha)L_T, \qquad L_T = \sum_{n=1}^{T} S_n. \tag{17}$$

For $T \geq 1$, let $p = N/T$ and use $0 \log 0 = 0$. Define the exact mixture redundancy

$$R_T(N) = N\log p + (T-N)\log(1-p) - \log \mathrm{B}(N+1/2, T-N+1/2) + \log\pi. \tag{18}$$

Write $\mathrm{KL}(p\|\alpha) = p\log(p/\alpha) + (1-p)\log[(1-p)/(1-\alpha)]$ for binary Kullback–Leibler divergence. Taking logarithms in Equation 16 proves

$$\begin{aligned} T\,\mathrm{KL}(p\|\alpha) &= \log(W_T/\epsilon) + R_T(N) \\ &\leq H_T := \log\{1 + (1-\alpha)L_T/\epsilon\} + R_T(N). \end{aligned} \tag{19}$$

All logarithms are natural. The identity holds also at $N = 0$ and $N = T$.

Lemma 1 (Elementary Jeffreys redundancy bound)**.** *For all $T \geq 1$ and $0 \leq N \leq T$, $0 \leq R_T(N) \leq \log\{\pi(T+1)/2\}$.*

Proof. The Bernoulli likelihood $x^N(1-x)^{T-N}$ has maximum $p^N(1-p)^{T-N}$. Its average under a probability density is no larger, which proves the lower bound on R_T. On $(0,1)$ the Jeffreys density satisfies $[\pi\sqrt{x(1-x)}]^{-1} \geq 2/\pi$. Its mixture likelihood is therefore at least

$$\frac{2}{\pi}\int_0^1 x^N(1-x)^{T-N}\,dx = \frac{2}{\pi(T+1)\binom{T}{N}}.$$

The binomial mass $\binom{T}{N}p^N(1-p)^{T-N}$ is at most one. The ratio of maximum likelihood to mixture likelihood is consequently at most $\pi(T+1)/2$. Taking logarithms completes the proof. □

For completeness, binary Pinsker follows by applying convexity to $D(p) = \mathrm{KL}(p\|\alpha) - 2(p-\alpha)^2$. In the interior, $D''(p) = 1/[p(1-p)] - 4 \geq 0$, while $D(\alpha) = D'(\alpha) = 0$. Continuity gives the endpoints, so $\mathrm{KL}(p\|\alpha) \geq 2(p-\alpha)^2$. Combining this fact with Equation 19 proves the sharper certificate

$$|p - \alpha| \leq b_T := \sqrt{H_T/(2T)}. \tag{20}$$

Lemma 1 then proves Proposition 1. If desired, intersect this bound with the trivial discrepancy bound $\max(\alpha, 1-\alpha)$.

If $S_n \leq Dn^r$ for fixed $D > 0$ and $r \geq 0$, then $L_T \leq D(T+1)^{r+1}/(r+1)$ by comparison with an integral. For fixed ϵ and α, Equation 19 is then $O(\log T)$ and $b_T = O(\sqrt{\log T/T})$. The envelope is a condition on the realized score sequence; the algorithm does not need D or r as an input. The finite certificate itself needs only finite scores.

B.3 A WORKED UPDATE AND NONNEGATIVE ENVELOPES

With $\alpha = .1$ and $\epsilon = .01$, the following decimal-return example checks the ordering of the native update:

Table 4: Worked signed-wealth updates in decimal-return units.

n	S_n	λ_n	Issued q_n	m_n	W_n
1	.010	.50	.044444	0	.005556
2	.050	.25	.009259	1	.013889
3	.002	.50	.061728	0	.007716

Values are rounded for display. The count used in λ_n precedes the current outcome. The second observation is evaluated against its issued radius, not the larger radius produced by its own miss.

Let C_n be the signed set and let $C_n^+ = [\mu_n - \max(q_n, 0), \mu_n + \max(q_n, 0)]$. Pointwise, $C_n \subseteq C_n^+$, so the envelope miss count is no larger than the signed miss count. On the same scored prefix its empirical miscoverage is at most $\alpha + b_T$. There is no corresponding lower bound around alpha from inclusion alone. For an all-zero score stream, any closed nonnegative interval covers every point; exact positive target matching instead requires the signed empty-set events. Both coverages can be recorded, but the wealth update always uses the signed event.

C DELAYED FEEDBACK AND EVALUATION WINDOWS

C.1 A CAUSAL THREE-LANE SCHEDULE

Let t count successive sessions in one market. Issue t has fixed lane $\ell(t) = t \bmod 3$, regardless of whether its future label is observed. The event sequence for one such issue is:

Table 5: Release and issue order for a single prediction.

Session/event	Action
t, after close	Freeze μ_{tf}, q_{tf}, a_{tf}, and stock scores.
$t+1$, close	Execute the stored stock decision.
$t+2$, close before trade	Mark its outcome and determine completeness.
$t+3$, release	Score the stored prediction and update its lane.
$t+3$, after release	Issue the next prediction in the same lane.

Release and issue on the same session are distinct ordered events. The economic mark satisfies $t+2 < t+3$; the release event must also precede issue. A replay assumes the completed daily bar is available by that release event. An unresolved label cannot be fed to a new prediction merely because its planned mark date has passed.

Proposition 2 (Reduction to native feedback)**.** *Suppose that each lane has at most one unresolved issue, stores its forecast and signed threshold, and resolves that issue before issuing again. A resolved missing label changes no learner state. After deleting these skipped issues, each lane follows*

the signed native recurrence. If nonempty lane l has T_l scored observations, N_l misses, and a discrepancy bound b_l, then

$$\left|\frac{\sum_l N_l}{\sum_l T_l} - \alpha\right| \le \frac{\sum_l T_l b_l}{\sum_l T_l}. \tag{21}$$

Proof. Before a lane's first observed issue, its state is the initialized state; any skipped issues have left it unchanged. Before each subsequent observed issue, the preceding observed outcome has been processed, and no other outcome has updated that lane. Its issued threshold is therefore the next action of the native recurrence. At release the residual uses the stored mean; the lane's radius has not changed in the meantime. Induction gives the native game on the observed subsequence. Shared mean updates from other lanes only change the next predictable score center and do not change this argument. Finally,

$$\left|\sum_l (N_l - \alpha T_l)\right| \le \sum_l |N_l - \alpha T_l| \le \sum_l T_l b_l.$$

Divide by the total observed count. No independence is used. □

Missingness is resolved under the same declared rule for all interval methods. An unavailable label is not a zero residual. It advances neither scored time, miss count, wealth, nor mean; it still consumes its issue slot. A later value cannot be inserted retrospectively after that lane has advanced. At the end of the financial horizon, complete pending labels may be processed in their lane order for an audit without issuing further allocations. A certificate formed before that flush concerns only the already scored prefix.

For two replays with identical information before an issue, the stored means, lane states, coefficients, and stock decisions at that issue coincide. To see this, induct over release and issue events: the same released label produces the same deterministic update, and the same issue inputs produce the same decision. Perturbing later labels or their completeness cannot alter an earlier action. This is an information-flow property of the replay, not an independence assumption about returns.

C.2 The exact continued-window bound

A calibrator carried from warmup is not a fresh learner at the beginning of a test year. Fix one lane and suppose a contiguous scored evaluation segment begins after m observations, k_0 misses, and wealth $W_m > 0$. Define

$$a = k_0 + 1/2, \qquad b = m - k_0 + 1/2. \tag{22}$$

For the following $n \ge 1$ observations, let k be the miss count and L the score sum. The sequential predictive fractions start from the beta prior (a, b), so the same multiplication used in Equation 15 gives

$$\frac{W_{m+n}}{W_m} = \frac{\mathrm{B}(a+k, b+n-k)}{\mathrm{B}(a,b)} \alpha^{-k} (1-\alpha)^{-(n-k)}. \tag{23}$$

Define

$$R_n^{a,b}(k) = k \log(k/n) + (n-k)\log(1-k/n) - \log \mathrm{B}(a+k, b+n-k) + \log \mathrm{B}(a,b). \tag{24}$$

The prior's integrated likelihood is no larger than the maximum likelihood, so $R_n^{a,b}(k) \ge 0$. Summing the one-step wealth bound only over this segment yields $W_{m+n} \le W_m + (1-\alpha)L$. Hence

$$n\,\mathrm{KL}(k/n \| \alpha) \le H_{n|m} := \log\{1 + (1-\alpha)L/W_m\} + R_n^{a,b}(k), \tag{25}$$

$$|k/n - \alpha| \le \sqrt{H_{n|m}/(2n)}. \tag{26}$$

These formulas prove the continued-window certificate directly. The posterior prior can be concentrated; its redundancy is not bounded by substituting n into the fresh-Jeffreys bound in Lemma 1.

For example, start at alpha .1 and force 30 warmup misses by selecting each finite score above its current threshold. Then supply 250 zero scores. Throughout those 250 observations the miss count stays 30 and $\lambda_n \ge 30.5/280 > .1$, so every test radius is positive and every zero score is covered. The test discrepancy is .1. Treating this as a fresh 250-observation zero-score sequence

would give the exact fresh-mixture radius approximately .08165, which is false for this continued state. Equation 26 retains the warmup prior and wealth and does not make that substitution.

For a reported window, save (m, k_0, W_m) separately for each factor and lane after its preceding feedback and before its first included prediction. A readiness rule based on a monotone observed count places earlier unready updates in m; it does not erase them. Pool the resulting bounds by each lane's included count using Equation 21. For an arbitrary subset B of a scored prefix of length T, with complement A, let $G_D = \sum_{j \in D}(\alpha - m_j)$. An alternative exact accounting is

$$\frac{|G_B|}{|B|} \le \frac{Tb_T + |G_A|}{|B|}, \qquad |B| > 0. \tag{27}$$

It follows from $G_B = G_{A \cup B} - G_A$. This formula handles excluded scored observations inside a reporting interval, for which a contiguous-window prior formula cannot simply be applied. Empty lanes or factors have undefined coverage and zero pooling weight. Macro-average factor coverage and pooled observation coverage use different denominators and are reported as such.

D Allocation, Group Controls, and Execution Geometry

D.1 Coefficient map and comparator definitions

Let J_t be the ready-factor set and $M_t = |J_t|$. For $f \in J_t$, write

$$r_{tf} = \frac{\#\{g \in J_t : u_{tg} < u_{tf}\} + (\#\{g \in J_t : u_{tg} = u_{tf}\} + 1)/2}{M_t}. \tag{28}$$

This is the average ordinal rank divided by M_t. Set $a_{tf} = \exp\{2(r_{tf} - 1/2)\}$ for ready factors and one otherwise. For $M_t > 0$, $1/M_t \le r_{tf} \le 1$, giving $e^{-1} < a_{tf} \le e$ and a ratio below e^2 between any two coefficients. If no factor is ready, every coefficient is one. If all ready utilities tie, their common multiplier is e^{1/M_t}; it cancels when every available factor has that multiplier but can change the relative contribution of unready factors.

With two ready utilities $u_1 < u_2$, ranks are $(1/2, 1)$ and coefficients are $(1, e)$. Suppose their stock scores are $(.9, .2)$ for stock A and $(.3, .8)$ for stock B. Equal coefficients tie the two stocks at .55. Dynamic scores are $(.9 + .2e)/(1 + e)$ and $(.3 + .8e)/(1 + e)$, so B ranks above A. This example changes stock selection using only an issue-time factor weighting; its scores do not specify any future return.

The following table gives the complete utility definitions. Every non-Fixed arm uses Equation 28 and the same observed-label mean.

Table 6: Allocation utilities before the common coefficient map.

Method	Utility before ranking
Fixed	Coefficients identically one
Mean	μ_{tf}
Rolling	$\mu_{tf} - \frac{1}{4}\max(q^{\mathrm{R}}_{tf}, 0)$
KT	$\mu_{tf} - \frac{1}{4}\max(q^{\mathrm{KT}}_{tf}, 0)$
UP	$\mu_{tf} - \frac{1}{4}\max(q^{\mathrm{UP}}_{tf}, 0)$
UP ratio	$\mu_{tf} / \max(q^{\mathrm{UP}}_{tf}, 10^{-4})$

Rolling sorts the latest at most 60 released scores across all three lanes. For n stored scores it uses order statistic $1 + \lceil .9(n-1) \rceil$, the `higher` empirical-quantile convention, and radius .01 if empty. The common readiness threshold is 60 complete labels. KT has three lane states, each initialized at $W_0 = .01$, $\beta_1 = q_1 = 0$, and uses

$$g_n = \alpha - \mathbf{1}\{S_n > q_n\}, \quad W_n = W_{n-1} - g_n q_n, \quad \beta_{n+1} = \frac{n\beta_n - g_n}{n+1}, \quad q_{n+1} = W_n \beta_{n+1}. \tag{29}$$

Here g_n is the signed coverage error and β_n the betting fraction. This is the symmetric betting comparator of Podkopaev et al. (2024). It differs from the asymmetric Jeffreys predictive fractions used by UP. The proof in Appendix B applies to the specified UP recurrence, rather than to every comparator by its name.

D.2 Complete Family definition

The six groups below are a fixed grouping by feature meaning, not an official Qlib taxonomy. Each rolling operator contributes all five of its lookbacks to the same group. The displayed counts refer to the 157-feature starting inventory; training support can reduce the retained count in a group.

Table 7: Fixed feature groups used by the Family combiner.

Group	Operators or feature block	Count
Price geometry	Nine candle features; OPEN0, HIGH0, LOW0	12
Trend and momentum	ROC, MA, BETA, CNTP, CNTN, CNTD, SUMP, SUMN, SUMD	45
Price position	MAX, MIN, QTLU, QTLD, RANK, RSV, IMAX, IMIN, IMXD	45
Dispersion and trend fit	STD, RSQR, RESI, WVMA	20
Volume dynamics	VMA, VSTD, VSUMP, VSUMN, VSUMD	25
Price–volume relation	CORR, CORD	10

The nine candle names are KMID, KLEN, KMID2, KUP, KUP2, KLOW, KLOW2, KSFT, and KSFT2. The groups are fixed before performance is computed.

Suppress t for one stock i. Let $\mathcal{G}_g$ be the retained factors in group g, $F_g = |\mathcal{G}_g|$, and O_i the available factors. Define

$$n_{ig} = |O_i \cap \mathcal{G}_g|, \qquad A_{ig} = \sum_{f \in O_i \cap \mathcal{G}_g} a_f, \tag{30}$$

$$m^a_{ig} = \frac{\sum_{f \in O_i \cap \mathcal{G}_g} a_f z_{if}}{A_{ig}}, \qquad m^1_{ig} = \frac{\sum_{f \in O_i \cap \mathcal{G}_g} z_{if}}{n_{ig}}. \tag{31}$$

Means are defined only for $n_{ig} > 0$; positive coefficients then give $A_{ig} > 0$. Family admits groups $E_i = \{g : n_{ig} \geq \max(1, \lceil .2F_g \rceil)\}$ and computes

$$s_i^{\text{Family}}(a) = \frac{1}{|E_i|} \sum_{g \in E_i} m^a_{ig}, \qquad |E_i| \geq \min(2, G), \tag{32}$$

where G counts retained nonempty groups. Otherwise its score is unavailable. This rule is applied after the common 80% stock-level feature requirement. Multiplying all coefficients in a group by one positive constant leaves m^a_{ig} and the Family score unchanged. On a complete inventory, factor f in group g receives effective weight $a_f/(G\sum_{j \in \mathcal{G}_g} a_j)$; at unit coefficients this is $1/(GF_g)$.

The Family-rebalanced control instead weights the eligible group means by $\bar{a}_g = F_g^{-1} \sum_{f \in \mathcal{G}_g} a_f$:

$$s_i^{\text{Family-rebalanced}}(a) = \frac{\sum_{g \in E_i} \bar{a}_g m^a_{ig}}{\sum_{g \in E_i} \bar{a}_g}. \tag{33}$$

This control restores a changing group weight to an equal-group baseline. The mean $\bar{a}_g$ uses the fixed group inventory; its definition is separate from the stock-specific All decompositions below. Standalone-group Fixed controls average that group's available oriented features with unit coefficients and use the same executor and group support requirement.

D.3 Exact within- and between-group controls with missing features

For the All controls, sum over $H_i = \{g : n_{ig} > 0\}$ and let $n_i = \sum_g n_{ig}$ and $A_i = \sum_g A_{ig}$. Their scores share the same stock-level availability rule; no additional Family eligibility test is inserted. The definitions are

$$s_i^0 = \sum_{g \in H_i} \frac{n_{ig}}{n_i} m^1_{ig}, \qquad s_i^{\text{All}} = \sum_{g \in H_i} \frac{A_{ig}}{A_i} m^a_{ig}, \tag{34}$$

$$s_i^{\text{Within}} = \sum_{g \in H_i} \frac{n_{ig}}{n_i} m^a_{ig}, \qquad s_i^{\text{Between}} = \sum_{g \in H_i} \frac{A_{ig}}{A_i} m^1_{ig}. \tag{35}$$

Thus Within changes relative coefficients inside each available group while preserving the baseline mass n_{ig}/n_i. Between changes group mass while keeping each available group internally equal-weighted. Full All changes both. These statements hold for each stock's actual missing-feature pattern.

Let $\delta\pi_{ig} = A_{ig}/A_i - n_{ig}/n_i$ and $\delta m_{ig} = m^a_{ig} - m^1_{ig}$. Expanding the product gives the exact score decomposition

$$s_i^{\mathrm{All}} - s_i^0 = \underbrace{\sum_g (n_{ig}/n_i)\delta m_{ig}}_{s_i^{\mathrm{Within}} - s_i^0} + \underbrace{\sum_g \delta\pi_{ig} m^1_{ig}}_{s_i^{\mathrm{Between}} - s_i^0} + \sum_g \delta\pi_{ig}\delta m_{ig}. \tag{36}$$

The final term is an interaction. Consequently, the full score change need not equal the sum of the two separate controls, and nonlinear stock selection does not make their portfolio returns additive.

Replacing a_f globally by $a_f/\bar{a}_g$ preserves group mass only on a fully available inventory. For a simple algebraic example, take two groups with two factors each and coefficients $(1, 2)$ and $(1, 1)$. If the first group's coefficient-one factor is missing for a stock, available counts are $(1, 2)$ and masses are $(2, 2)$. The correct Within mass of the first group is $1/3$. Global division by the full-group mean 1.5 instead gives mass $(2/1.5)/(2/1.5 + 2) = .4$. Equations 35 avoid this discrepancy by using the stock's available factors.

D.4 Linear intervals and the execution discrepancy

Let Y_f be the simultaneous standalone factor returns, and let w and v be two nonnegative vectors summing to one. For nonnegative radii $q_f^+ = \max(q_f, 0)$, on the event $|Y_f - \mu_f| \le q_f^+$ for every factor,

$$(w - v)^\top Y \ge (w - v)^\top \mu - \sum_f |w_f - v_f| q_f^+. \tag{37}$$

Indeed, each term $(w_f - v_f)(Y_f - \mu_f)$ is at least $-|w_f - v_f|\,|Y_f - \mu_f|$, and summation proves the result. This is a deterministic bound on a simultaneous event. Marginal temporal coverage for each factor does not supply that event at a chosen trading date.

For a coefficient trajectory $a_{1:t}$, let $V_t[a_{1:t}]$ denote the net return of the actual combined-stock ledger. Define diagnostic inventory weights $w_{tf} = a_{tf}/\sum_j a_{tj}$ and

$$\kappa_t[a_{1:t}] = V_t[a_{1:t}] - w_t^\top Y_t. \tag{38}$$

These inventory weights describe a linear comparator; actual stock-score weights additionally depend on available features. For another coefficient trajectory b and diagnostic weights v, subtraction yields

$$V_t[a_{1:t}] - V_t[b_{1:t}] = (w_t - v_t)^\top Y_t + \kappa_t[a_{1:t}] - \kappa_t[b_{1:t}]. \tag{39}$$

The discrepancy includes ranking, position persistence, feature availability, and transaction costs. Equations 37 and 39 separate the interval calculation from the executed comparison. No lower bound on the difference in κ is assumed.

E Assumptions and the Populations Being Measured

The proof uses fixed $\alpha \in (0, 1)$, positive initial wealth, finite nonnegative absolute scores, and the exact signed update on the observed sequence. The asymptotic conclusion additionally uses a polynomial score envelope. The calculation requires neither exchangeability nor independent returns. Finite precision is checked by positive finite wealth and finite radii; numerical failure is not repaired by an unreported reset, cap, or target change.

The delayed extension assumes deterministic lane assignment at issue and resolution before lane reuse. A shared mean is allowed because the score center is predictable and stored. The completed-label observation mask defines the population being calibrated. The proof gives no statement about

unobserved marks, a subset selected by allocation, or a simultaneous cross-factor event. The full-prefix and continued-window formulas specify exactly which scored observations are counted.

The financial replay assumes the published archive's session alignment, adjustment convention, membership intervals, and availability contract. Feature construction is causal with respect to those supplied inputs. A daily-bar benchmark uses its declared close-price fills, budget sizing, fee schedule, and missing-mark accounting. Those execution assumptions are shared across comparators and distinct from the mathematical score-sequence assumptions. The five-stage chain is tested financially by running its entire holdings path; coverage, forecast error, and executed utility remain separate measured quantities.

F Fractional Feedback, Delayed Release, and Missing Outcomes

This appendix proves Proposition 3 for the specified policy-derived weights. The Jeffreys mixture, signed bet, and Pinsker step are inherited from universal-portfolio OCP (Liu et al., 2026). Soft-group portfolio OCP is also established (Bharti et al., 2026); the claim here is the precise multi-label feedback unit and its observed-population identification.

F.1 Construction and finite-sample certificate

Factorwise coverage does not describe the factors that matter most to the current stock ranking. We evaluate an influence-weighted universal-portfolio extension (IUP), where influence means local stock-score sensitivity. At issue, let a^{μ}_{tf} be the Mean coefficient and $p_{tif} = a^{\mu}_{tf} / \sum_{g \in O_{ti}} a^{\mu}_{tg}$ on a stock's available factors. Evaluating the score at a^{μ}_{t}, differentiation of Equation 7 gives

$$\frac{\partial s_{ti}}{\partial \log a_f} = p_{tif}(z_{tif} - s_{ti}). \tag{40}$$

Average the absolute derivative over eligible stocks and normalize across factors to obtain the issue-time weights $v_{tf} \geq 0$, $\sum_f v_{tf} = 1$. Unavailable factors have zero derivative. If all averaged derivatives vanish, set $v_{tf} = 1/F$. Sensitivity uses All for both combiners; it is not a factor's marginal contribution to executed return.

Let σ_{tf} be an issue-time exponentially weighted moving average (EWMA) of absolute errors, initialized at .01, updated on released complete labels with half-life 20 and floored at 10^{-4}. The standardized score is $X_{tf} = |Y_{tf} - \mu_{tf}|/\sigma_{tf}$. If O_t contains the complete released factor labels, define observed sensitivity mass $H_t = \sum_{f \in O_t} v_{tf}$ and normalized observed weights $w_{tf} = v_{tf}/H_t$ for $f \in O_t$. Skip feedback when $H_t = 0$. With a common signed radius q_t, the fractional miss is

$$m_t = \sum_{f \in O_t} w_{tf} \mathbf{1}\{X_{tf} > q_t\} \in [0, 1]. \tag{41}$$

The same three release lanes apply. In the next recurrence, t indexes nonempty observed updates within one lane. Each constant fraction $\lambda \in (0, 1)$ has wealth $A_t(\lambda)$ and bet $b(\lambda) = (\lambda - \alpha)/[\alpha(1 - \alpha)]$:

$$A_t(\lambda) = A_{t-1}(\lambda)[1 + (m_t - \alpha)b(\lambda)], \quad q_t = \int A_{t-1}(\lambda) b(\lambda)\, \pi(d\lambda). \tag{42}$$

Initialize $A_0(\lambda) = \epsilon = .01$. The Jeffreys prior is the probability measure $\pi(d\lambda) = d\lambda/[\pi\sqrt{\lambda(1-\lambda)}]$; $W_t = \int A_t\, d\pi$ is total mixture wealth. The financial coefficient utility becomes $\mu_{tf} - .25\sigma_{tf} \max(q_t, 0)$; its ranking map, stock-score combiner, and Top50/Drop5 ledger do not change. Uniform-weighted universal-portfolio feedback (UUP) is the ablation with $v_{tf} = 1/F$ before observed-label normalization.

Proposition 3 (Fractional-wealth certificate and missingness)**.** *For one lane of T nonempty observed updates, write $M = \sum_{t=1}^{T} m_t$, $p = M/T$, and $L = \sum_t \sum_{f \in O_t} w_{tf} X_{tf}$. Define binary Kullback–Leibler divergence as $\mathrm{KL}(p\|\alpha) = p\log(p/\alpha) + (1-p)\log[(1-p)/(1-\alpha)]$, using natural logarithms and $0 \log 0 = 0$. For finite nonnegative scores, $W_T > 0$ and*

$$T\,\mathrm{KL}(p\|\alpha) \leq \log(1 + L/\epsilon) + R_T(M). \tag{43}$$

Here $R_T(M)$ is the log ratio of the best constant Bernoulli likelihood to its Jeffreys mixture; Appendix F gives the exact beta-function formula, including noninteger M. For observed mass $H_t > 0$, the full-factor miss at issue t lies in $[H_t m_t,\ H_t m_t + 1 - H_t]$ without an assumption on missing labels. For a nonnegative radius, both endpoints are attainable by nonnegative hidden scores. If $q_t < 0$, every score misses and the full-factor miss is exactly one.

F.2 Proof and identification

Issue-time sensitivity. For an eligible stock, put $D_i = \sum_{f \in O_i} a_f$ and $p_{if} = a_f / D_i$. Holding its availability and factor scores fixed, $\partial p_{ig} / \partial \log a_f = p_{ig}(\mathbf{1}\{f = g\} - p_{if})$. Therefore

$$\frac{\partial s_i}{\partial \log a_f} = \sum_{g \in O_i} z_{ig} \frac{\partial p_{ig}}{\partial \log a_f} = p_{if}(z_{if} - s_i).$$

Set this derivative to zero for an unavailable factor, average its absolute value over all currently eligible stocks, and normalize; this constructs v_{tf} before outcomes. It is a local derivative at the Mean coefficients. The subsequent percentile-rank coefficient map and Top50/Drop5 operator are discontinuous at ties and selection boundaries, so the derivative is not a bound on executed return changes.

Exact fractional mixture. For one lane let $\pi(d\lambda) = [\pi\sqrt{\lambda(1-\lambda)}]^{-1} d\lambda$. With $A_0(\lambda) = \epsilon$, define

$$b(\lambda) = \frac{\lambda - \alpha}{\alpha(1-\alpha)}, \qquad A_t(\lambda) = \epsilon \prod_{j=1}^{t} [1 + (m_j - \alpha) b(\lambda)], \tag{44}$$

$$W_t = \int_0^1 A_t(\lambda)\pi(d\lambda), \qquad q_t = \int_0^1 A_{t-1}(\lambda) b(\lambda) \pi(d\lambda). \tag{45}$$

The multiplier is the convex combination

$$1 + (m - \alpha) b(\lambda) = m\frac{\lambda}{\alpha} + (1 - m)\frac{1 - \lambda}{1 - \alpha} > 0 \tag{46}$$

for $0 < \lambda < 1$ and $0 \le m \le 1$. Thus $W_t > 0$ and integration yields $W_t - W_{t-1} = (m_t - \alpha) q_t$ exactly. At $m = 0$ or 1 this is the ordinary two-asset factor; intermediate m is a mixture of returns, not a fictional binary coin toss.

The weighted arithmetic–geometric mean inequality applied to Equation 46 gives

$$1 + (m - \alpha) b(\lambda) \ge (\lambda/\alpha)^m ((1 - \lambda)/(1 - \alpha))^{1 - m}.$$

Multiplying and integrating, with $M = \sum_{t=1}^{T} m_t$, proves

$$\frac{W_T}{\epsilon} \ge \frac{\mathrm{B}(M + 1/2, T - M + 1/2)}{\pi \alpha^M (1 - \alpha)^{T - M}}. \tag{47}$$

For $q_t < 0$, every nonnegative X_{tf} exceeds q_t, so $m_t = 1$ and $(m_t - \alpha) q_t < 0$. For $q_t \ge 0$ and $m_t \le \alpha$ the change is nonpositive. Otherwise

$$(m_t - \alpha) q_t \le m_t q_t \le \sum_{f \in O_t} w_{tf} X_{tf},$$

because each missed score exceeds q_t. Summation gives $W_T \le \epsilon + L$. This proof allows weights to be renormalized after the observation mask is known: it is pathwise on that realized observed subsequence, not a promise about the missing factors.

Let $p = M/T$ and use $0 \log 0 = 0$. The real-count Jeffreys redundancy is

$$R_T(M) = M \log p + (T - M) \log(1 - p) - \log \frac{\mathrm{B}(M + 1/2, T - M + 1/2)}{\pi}. \tag{48}$$

The beta integral is no larger than the maximum of $\lambda^M (1 - \lambda)^{T - M}$, so $R_T(M) \ge 0$. Taking logarithms in the lower and upper bounds and rearranging proves the proposition. Binary Pinsker supplies the auxiliary discrepancy bound

$$|p - \alpha| \le \sqrt{\{\log(1 + L/\epsilon) + R_T(M)\}/(2T)}. \tag{49}$$

The beta formula permits real M, including $M = 0, T$. It is a finite data-dependent certificate; no independent-sampling assumption enters.

Exact quadrature and delayed lanes. The degree-t product $A_t(\lambda)$ is a polynomial in λ. Under the arcsine transformation $\lambda = (1 + \cos\theta)/2$, K midpoint Chebyshev nodes integrate degree at most $2K - 1$ exactly. The implementation uses $K = 1024$ and a maximum of 551 updates in a lane, below that threshold. Numerical sums remain subject to floating-point roundoff; positivity and wealth accounting are tested. In three lanes, feedback from issue t is processed at issue $t + 3$ before the next issue in its lane. Issue-time v, μ, σ, q are frozen, and a missing label does not advance wealth. The pathwise result applies separately to each scored lane. It does not include a terminal issue whose outcome has not yet been released.

Warmup and a continued window. After s scored updates define the exact wealth posterior $\pi_s(d\lambda) = A_s(\lambda)\pi(d\lambda)/W_s$. For a following contiguous window of n scored updates, $M' = \sum m_t$, $p' = M'/n$, and $L' = \sum_t \sum_{f \in O_t} w_{tf} X_{tf}$, set

$$R_n^{\pi_s}(M') = M' \log p' + (n - M') \log(1 - p') - \log \int \lambda^{M'} (1 - \lambda)^{n - M'} \pi_s(d\lambda). \tag{50}$$

The same arithmetic–geometric mean argument, now integrated against π_s, and $W_{s+n} \le W_s + L'$ prove

$$n \,\mathrm{KL}(p' \| \alpha) \le \log(1 + L'/W_s) + R_n^{\pi_s}(M'). \tag{51}$$

Unlike the binary UP posterior, π_s is generally not a beta density: fractional returns are arithmetic mixtures. Substituting beta pseudo-counts into Equation 51 would be incorrect. All empirical test-window rates are therefore descriptive; the computed finite certificate uses each lane's complete scored prefix.

Identification with incomplete factor labels. Let v_{tf} sum to one over all issued factors, $H_t = \sum_{f \in O_t} v_{tf}$, and m_t be the conditional observed miss for $H_t > 0$. The full miss is

$$m_t^{\mathrm{full}} = H_t m_t + \sum_{f \notin O_t} v_{tf} \mathbf{1}\{X_{tf} > q_t\}.$$

Each hidden indicator is in $[0, 1]$, proving $H_t m_t \le m_t^{\mathrm{full}} \le H_t m_t + 1 - H_t$. When $q_t \ge 0$, all hidden scores zero or all above q_t attain the respective endpoints without changing any observed update. When $q_t < 0$, nonnegative scores force every hidden indicator to one, so $m_t^{\mathrm{full}} = 1$. For $H_t = 0$, $[0, 1]$ is a valid bound, reduced to $\{1\}$ for a negative radius. Averaging the reported endpoints gives a valid outer interval for the full-factor mean; it can be tightened using negative-radius issues. This is an identification statement, not a confidence interval. Without a lower bound on H_t or an assumption on missing outcomes, observed conformal calibration cannot identify full-factor calibration.

G Public Data and Feature Construction

G.1 Sources, markets, and historical universes

We use daily China A-share and US equity archives in Qlib format. The Qlib README identifies Yahoo Finance as the source of its example market data and notes quality limitations. Its download implementation points to the SunsetWolf archive mirror. Both references are pinned to commit `be725493eb1a6bbb42bf11b37aa7669f59610ff1`. The selected releases are China v1 and US v2, acquired on September 16, 2026. Acquisition records retain archive sizes, SHA-256 identities, and member manifests. Qlib supplies the documented collection route and format; SunsetWolf supplies the downloaded copies. The records do not independently establish an unbroken vendor-to-mirror lineage or the correctness of every corporate-action adjustment.

The stock pools use the archives' CSI300 and S&P 500 membership intervals. For stock i, membership at session t is the union of its supplied inclusive intervals. Thus a stock can enter, leave, and re-enter the candidate universe. The table describes the loaded panels, including their small calendar buffer after the strategy endpoint. Symbol counts are unions over this panel period, not the number tradable on a particular date.

Table 8: Public archive panels and observed data availability.

Panel property	China	US
First session	2007-01-04	2007-01-03
Last loaded session	2020-08-07	2020-08-07
Loaded sessions	3,257	3,424
Historical symbols	690	733
Symbols lacking an archive directory	7	82
Finite closes within membership	95.154%	85.479%
Retained feature columns	157	157

Finite-close availability is the number of finite close entries divided by the number of expected stock-session entries within membership intervals; missing symbols remain in this denominator. The complete S&P 500 membership file contains 745 distinct symbols, of which 733 overlap the loaded period. Supplied intervals and historical removals do not certify a survivorship-free universe. Delisting coverage, revision timing, and historical availability of adjustment factors remain source assumptions.

The source audit identifies 82 absolute adjacent-session adjusted-price moves above 100% in the US 2008–2012 training period and none in 2013–2020; the China audit finds none in 2008–2020. This check requires observable adjacent prices and membership on the later session. It does not exhaust anomalies in prices used to value positions carried beyond membership or across quote gaps. No return is winsorized or removed because of its magnitude. Raw quote availability and the completeness of a strategy's realized labels are different quantities. The latter depends on its holdings and is reported with the corresponding calibration results.

Training-source sensitivity. A post-replay diagnostic rebuilds the original US training rank information coefficients (ICs), the cross-sectional rank correlations used to orient each factor, and then excludes only eligible stock-return pairs with absolute return above one, reranking the remaining observations. The original ICs, counts, active columns and orientations match the saved training metadata exactly. The filter removes 82 of 490,458 pairs (0.01672%) across 80 of 1,257 training dates. It changes two of 157 signs, `RSQR10` and `SUMP30`, whose mean ICs are near zero; no feature loses its training-support eligibility. The mean and maximum absolute changes in mean IC are 0.000105 and 0.000424. This diagnostic changes return-pair eligibility only: feature histories retain the original observations, so it is not a repaired-data replay. The fixed orientations and reported strategy results are retained. The per-feature diagnostic is included in the public result package.

G.2 Chronology and preprocessing

History begins in 2007 to initialize trailing features. Orientations use 2008–2012 only, with every forward label's final mark on or before December 31, 2012. The calendar year 2013 is reserved for development. Standalone and combined portfolios start their fixed-coefficient warmup in 2014; adaptive allocation is evaluated from January 1, 2015 through July 31, 2020. An issue must have its two-session mark within that endpoint. The loaded August buffer contributes neither later evaluation returns nor future labels to fitting. Each market uses its own supplied session calendar.

Adjusted open, high, low, close, and volume are read as supplied; the adjustment factor is used separately for share-unit conversion. Missing calendar rows are retained. Features are not forward-filled, backfilled, interpolated, or cross-sectionally imputed. Nonpositive current close or volume invalidates the bar for feature construction. Undefined or nonfinite feature outputs are unavailable. The archive contains no usable VWAP input for this specification, so `VWAP0` is omitted without constructing a proxy. No full-sample scaler or outcome-based feature filter is fitted.

For orientation, the stock label is $C_{t+2,i}/C_{t+1,i} - 1$, where $C_{t,i}$ is stock i's adjusted close at session t. On each training date, factor values and forward returns are converted to cross-sectional percentile ranks on their eligible observations; correlation is computed on paired available ranks. A supported feature-date requires at least 30 stocks and a finite correlation. At least 252 such dates are required per feature. The sign of the mean rank IC fixes orientation, with zero mapped to $+1$. All 157 columns pass this support requirement in both panels. Rank correlation is bounded in $[-1, 1]$: an extreme return affects ordering but cannot enter as an arbitrarily large numerical weight. This does

not repair a corrupted ordering. At issue, oriented features are ranked among available members with average ranks for ties; a stock needs at least $\lceil .8(157) \rceil = 126$ available features. The same mask is used by all allocation methods.

The orientation label fixes a feature's direction using stock returns. Online calibration instead targets the net return of its persistent standalone factor strategy, including cash, fees, turnover, and the preceding portfolio state.

G.3 Complete 157-column feature dictionary

The specification follows the pinned Alpha158 loader and its operators. Suppress stock and current-session subscripts. Write O, H, L, C, V for the current adjusted open, high, low, close, and volume and set $\delta = 10^{-12}$, $D = H - L$, $U = H - \max(O, C)$, $B = \min(O, C) - L$, and $Z = 2C - H - L$. The nine candle and three relative-price features occur once. Each of the 29 rolling prefixes below occurs at every $d \in \{5, 10, 20, 30, 60\}$, giving $9 + 3 + 29 \cdot 5 = 157$ columns. Names append d to the rolling prefix.

Let M_d, S_d, Σ_d, Q_d^p, and ρ_d denote trailing mean, sample standard deviation, sum, quantile, and Pearson correlation. $R_d(C)$ is the current close's average percentile rank within its window. I_d^H, I_d^L are the one-based positions of the first high maximum and low minimum in the actual window. Write b_d, R_d^2, e_d for the ordinary least-squares (OLS) slope, coefficient of determination, and endpoint residual from regressing close on session positions with an intercept. Also define $\Delta C = C - C_{-1}$, $\Delta V = V - V_{-1}$, $x = |C/C_{-1} - 1|V$, and $y^+ = \max(y, 0)$. Each comma-separated expression below corresponds to the name in the same position; trailing operators act on the entire indicated series.

Table 9: Complete feature formulas. Subscript $-d$ denotes a positional lag. The constants appear only in the denominators shown.

Name or rolling prefix	Formula
KMID, KLEN	$(C - O)/O$, D/O
KMID2	$(C - O)/(D + \delta)$
KUP, KUP2	U/O, $U/(D + \delta)$
KLOW, KLOW2	B/O, $B/(D + \delta)$
KSFT, KSFT2	Z/O, $Z/(D + \delta)$
OPEN0, HIGH0, LOW0	O/C, H/C, L/C
ROC, MA, STD	C_{-d}/C, $M_d(C)/C$, $S_d(C)/C$
BETA, RSQR, RESI	b_d/C, R_d^2, e_d/C
MAX, MIN	$\max_d(H)/C$, $\min_d(L)/C$
QTLU, QTLD	$Q_d^{.8}(C)/C$, $Q_d^{.2}(C)/C$
RANK	$R_d(C)$
RSV	$(C - \min_d(L))/(\max_d(H) - \min_d(L) + \delta)$
IMAX, IMIN, IMXD	I_d^H/d, I_d^L/d, $(I_d^H - I_d^L)/d$
CORR	$\rho_d(C, \log(V + 1))$
CORD	$\rho_d(C/C_{-1}, \log(V/V_{-1} + 1))$
CNTP, CNTN	$M_d(\mathbf{1}\{C > C_{-1}\})$, $M_d(\mathbf{1}\{C < C_{-1}\})$
CNTD	CNTPd minus CNTNd
SUMP, SUMN	$\Sigma_d((\Delta C)^+)/(\Sigma_d(\lvert\Delta C\rvert) + \delta)$, $\Sigma_d((-\Delta C)^+)/(\Sigma_d(\lvert\Delta C\rvert) + \delta)$
SUMD	SUMPd minus SUMNd
VMA, VSTD	$M_d(V)/(V + \delta)$, $S_d(V)/(V + \delta)$
WVMA	$S_d(x)/(M_d(x) + \delta)$
VSUMP, VSUMN	$\Sigma_d((\Delta V)^+)/(\Sigma_d(\lvert\Delta V\rvert) + \delta)$, $\Sigma_d((-\Delta V)^+)/(\Sigma_d(\lvert\Delta V\rvert) + \delta)$
VSUMD	VSUMPd minus VSUMNd

The fixed groups contain price geometry (12 columns), trend and momentum (45), price position (45), dispersion and trend fit (20), volume dynamics (25), and price–volume relations (10). All lookbacks of one operator stay together. These are study-defined economic groups; the complete assignment and aggregation rules appear in Appendix D.

Windows include the current session and use available observations from at most d rows, with a minimum of one observed value; a finite feature therefore does not imply a complete d-session history. Standard deviations use $n - 1$ degrees of freedom and quantiles use linear interpolation. Positional lags and OLS retain the spacing of missing sessions. OLS needs two observations; its residual is unavailable when the endpoint is missing. Qlib's 2×10^{-5} near-zero standard-deviation guard applies to R_d^2 and both marginal series in a correlation. Correlations use paired observations. The positional-extremum operators preserve the pinned implementation's first-occurrence rule, including its treatment of an internal not-a-number (NaN) missing value as an extremum index; an all-missing window is unavailable. Comparisons against NaN are false, so change-count features can remain finite across gaps. Stock-level feature support and execution tradability are checked separately. Source-level synthetic checks compare all feature names, missing masks, numerical values, and prefix invariance; they verify operator construction, not archive history.

G.4 Execution, costs, and adjustment-unit audit

Signals use information through close t, execute at close $t + 1$, and are marked at close $t + 2$ before the next trade. Completed factor labels enter estimation at $t + 3$, before its issue event. Their means and radii use stored forecasts, three separate feedback lanes, and the fixed parameters in Section 3. Missing quotes are carried only for portfolio accounting. Completeness requires that the larger of the missing-quote position values before entry and at the next mark, divided by pre-entry NAV, is at most 10^{-12}. This is a dimensionless accounting tolerance. Only finite complete labels update the estimators. When a quote returns, valuation includes the change from the carried mark. Positions remain accounted for after a stock leaves the signal universe until an allowed sale occurs.

Each portfolio starts with 10^8 local currency units, Chinese yuan (CNY) or US dollars (USD). The common Top50/Drop5 operator preserves its own holdings between rebalances and uses a 95% investment budget for new purchases. Fills require finite positive close, volume, and adjustment factor. China additionally requires an absolute adjacent-close change below 9.5%, the benchmark's fixed tradability convention; the US has no corresponding price-change filter. Buy and sell costs are respectively $\max(5, 0.0005v)$ and $\max(5, 0.0015v)$ for executed notional $v > 0$. An unfilled order has no fee. Cash earns zero; spread, slippage, and capacity effects are not modeled. Double-cost sensitivity multiplies both fees by two, holds the base-cost forecasts and coefficients fixed, and reruns the holdings ledger. There is no terminal liquidation. Net labels use pre-entry NAV as their denominator and include entry fees exactly once, as derived in Appendix A.

Let $P^{\mathrm{adj}} = bP^{\mathrm{raw}}$, $b > 0$. Raw shares equal bh^{adj}, so an execution lot is $100/b$ adjusted shares in China and $1/b$ in the US. With the stock scores fixed, a change of execution units by any constant $k > 0$ obeys

$$P^{\mathrm{adj}} \mapsto kP^{\mathrm{adj}}, \qquad b \mapsto kb, \qquad h^{\mathrm{adj}} \mapsto h^{\mathrm{adj}}/k. \tag{52}$$

It preserves raw prices, raw holdings, notionals, and cash flows. An absolute cutoff on adjusted shares violates this invariance: a numerically tiny share count can represent a large position. The executor identifies holdings by $h^{\mathrm{adj}} > 0$; a full sale assigns exactly zero. Its regression fixture rescales prices and adjustment factors by 10^{24} in both markets, then checks identical held sets, rescaled share quantities, economic positions, cash, and return records, including exits after scores become unavailable. This correction applies to all factor and combined ledgers under the same data, chronology, coefficient rules, and costs. It corrects inventory bookkeeping without changing or certifying the source observations.

H Complete Public Results and Aggregation Mechanisms

H.1 Architectures and presentation

The public protocol includes two score-combination architectures. All combines every available feature with its coefficient before selecting stocks. Family normalizes within each of six fixed feature groups and then averages the available group scores. Both use the same stock executor; their signal normalization gives different implicit group budgets. The main text presents All because its UP correction improves the pooled Fixed comparison in both markets. This is an expository choice after observing the predefined results, not a separately confirmed architecture selection. Every Family method, annual block, cost replay and mechanism control is retained here and in the accompanying machine-readable tables. No architecture is removed from the public study.

The word "better" depends on the comparison. Family's Fixed baseline earns 5.04 bp/day in China, above All Fixed's 2.16 bp/day; Family UP also has a higher absolute Chinese mean than All UP, at 4.07 versus 3.35 bp/day. Yet the correction reduces Family's own mean by 0.97 bp/day and increases maximum drawdown from 47.74% to 54.92%. In the US, Family UP improves its Fixed mean by 0.14 bp/day, but drawdown rises from 47.68% to 49.36%. Thus All gives the more consistent direction of the *correction relative to its own baseline*; it does not dominate Family in absolute return.

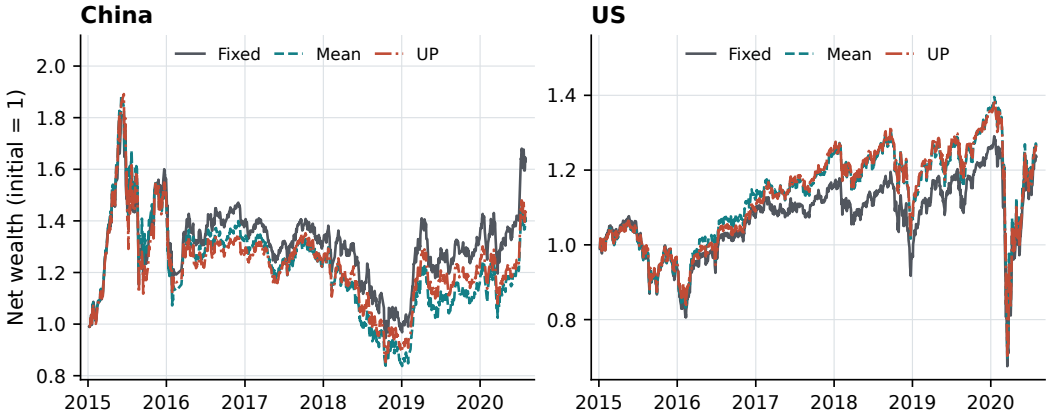

Figure 3: The complete Family trajectories for the two primary controls and UP, using the same public periods and base costs as the main text.

H.2 Testing a dominant-family explanation

Equation 7 assigns a fully observed group's total coefficient mass in proportion to the number and weights of its features. With unit coefficients, the two 45-feature groups each receive 28.66% of All's mass, whereas the 10-feature comovement group receives 6.37%. Family gives each available group an equal vote. This normalization changes the stock ranking even before uncertainty is added. The exact decomposition with missing feature support appears in Appendix D.2.

The frozen Within control changes the weighted score inside each group while retaining its baseline available-feature mass. Between changes group masses while retaining unweighted scores inside each group. In China, Within earns 2.40 bp/day and Between 3.33 bp/day, compared with 2.16 for Fixed and 3.35 for full UP. Most of the observed full UP improvement is reproduced by the between-group control. In the US, Within earns 2.35 bp/day, Between 2.80, and full UP 2.48, compared with 2.12 for Fixed. The effects are not additive: changing both components changes rankings, turnover and subsequent holdings.

The Rebalanced control restores the mean dynamic coefficient of each group after Family's within-group normalization. It earns 3.80 bp/day in China, below both Family UP and Family Fixed, and 2.67 bp/day in the US, close to Family UP. Restoring group mass therefore does not by itself repair

Table 10: Feature counts, full-support baseline mass and mean UP coefficient mass over the test period. These are signal coefficients, not realized portfolio exposures. Per-stock masses additionally depend on feature availability.

Group	Features	Fixed %	China UP %	US UP %
Dispersion / trend quality	20	12.74	12.64	13.88
Price geometry	12	7.64	8.50	8.38
Price position	45	28.66	30.67	28.00
Price–volume comovement	10	6.37	7.50	7.05
Trend / momentum	45	28.66	25.07	25.52
Volume dynamics	25	15.92	15.62	17.18

the Chinese Family correction. The outcome points to an interaction between group normalization, correlated ranks and the persistent stock portfolio.

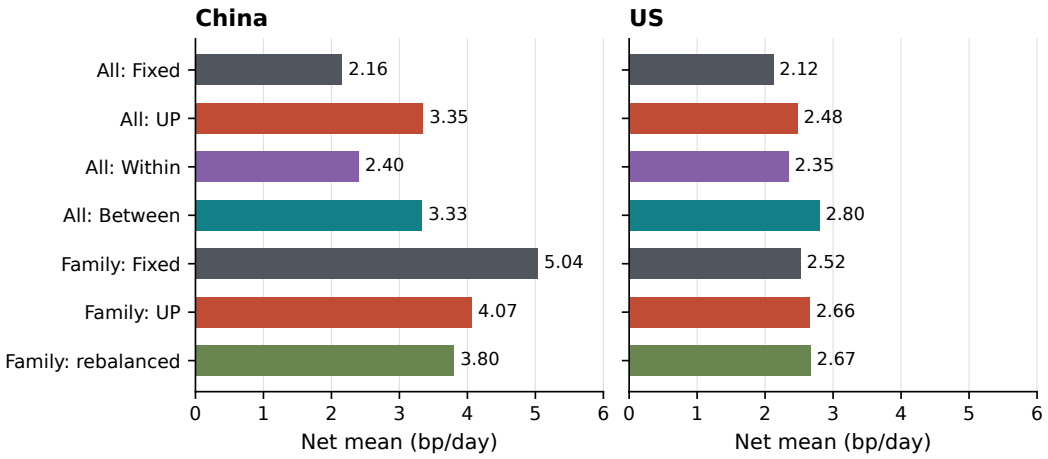

Figure 4: Predefined mechanism controls at base costs. All bars are full-period net means; Within retains baseline group masses and adapts within-group weights; Between adapts group masses and retains equal within-group weights. Family rebalanced restores dynamic group means to the Family combination.

The six standalone group strategies further constrain the proposed dominance explanation. Chinese comovement has the highest standalone mean, 4.94 bp/day, while volume dynamics has the lowest, 0.18. US trend/momentum leads at 3.22 bp/day and volume dynamics earns 2.01. The leading group changes by market, and none of these standalone returns equals that group's marginal contribution to a combined stock ranking. The evidence supports studying group budgets explicitly; it does not identify one universally dominant group or a single-family cause of every architecture difference.

H.3 Reading the complete comparisons

The tables retain every predefined method. F- denotes Family and G- a standalone feature-group strategy; Within, Between, and F-Rebalanced are defined in Appendix D. Each market has 21 combined trajectories at each cost setting and 157 standalone factor ledgers at base costs. Across two markets and seven reporting periods, these produce 588 strategy-period rows and 504 paired contrasts, not independent replications.

Net means are arithmetic; CAGR compounds and annualizes by 252 sessions. Drawdown includes initial wealth of one at each reporting boundary. For n days, $\mathrm{ES}_{.05}$ is the negative mean of the worst $\lceil .05n \rceil$ returns. Turnover counts buys and sells. The accompanying comma-separated values (CSV) files retain dates, gross means, volatility, costs, and accounting residuals.

H.4 CHINA: COMPLETE METHODS, 1X COSTS

All methods use the same market, chronology, factor signs, trade eligibility and execution rules. Net means and fees are in basis points per day; CAGR, maximum drawdown and expected shortfall are percentages.

Table 11: China: full-period performance at $1\times$ costs.

Method	Net bp/d	CAGR	Sharpe	MDD	$ES_{.05}$	Fee bp/d
Fixed	2.16	1.91	0.20	61.08	4.41	1.23
Mean	3.66	5.71	0.34	56.85	4.45	1.24
Rolling	3.19	4.57	0.30	56.52	4.41	1.12
KT	2.83	3.70	0.27	59.65	4.35	1.07
UP	3.35	5.14	0.32	53.21	4.32	1.10
UP-ratio	3.36	4.87	0.31	58.22	4.50	1.23
F-Fixed	5.04	9.81	0.49	47.74	4.18	1.01
F-Mean	3.92	6.67	0.38	55.63	4.26	1.02
F-Rolling	3.55	6.00	0.36	56.62	4.14	0.90
F-KT	4.06	7.14	0.40	54.41	4.27	0.97
F-UP	4.07	7.00	0.39	54.92	4.35	1.10
F-UP-ratio	4.00	6.89	0.39	57.59	4.36	1.02
Within	2.40	2.50	0.23	59.64	4.42	1.21
Between	3.33	5.01	0.32	55.00	4.34	1.22
F-Rebalanced	3.80	6.76	0.39	55.80	4.12	0.71
G-Dispersion	3.49	4.35	0.29	56.96	5.05	1.04
G-Geometry	3.39	6.04	0.37	42.06	3.91	1.37
G-Position	2.49	2.91	0.24	57.41	4.27	1.14
G-Comovement	4.94	10.12	0.53	43.87	3.91	0.79
G-Trend	3.04	4.27	0.29	54.76	4.35	1.31
G-Volume	0.18	-2.74	0.02	61.27	4.37	1.28

Gross traded notional, stock exposure and affected missing-mark weight are daily means in percent of pretrade NAV. Affected weight is the larger of missing-mark exposure before trading and at the label mark.

Table 12: China: turnover, exposure, and missing-mark weight at $1\times$ costs (percent).

Method	Turnover	Exposure	Affected weight
Fixed	12.31	94.79	3.96
Mean	12.41	94.79	3.85
Rolling	11.15	94.72	4.70
KT	10.65	94.82	5.04
UP	10.97	94.85	4.69
UP-ratio	12.33	94.75	3.92
F-Fixed	10.08	94.87	5.55
F-Mean	10.17	94.84	5.66
F-Rolling	9.00	94.90	6.42
F-KT	9.73	94.89	5.94
F-UP	10.98	94.88	4.86
F-UP-ratio	10.17	94.90	5.35
Within	12.14	94.79	4.06
Between	12.18	94.82	4.02
F-Rebalanced	7.13	94.92	7.19
G-Dispersion	10.35	94.84	5.40
G-Geometry	13.68	94.87	3.37
G-Position	11.41	94.80	4.66
G-Comovement	7.86	94.91	6.62
G-Trend	13.08	94.76	3.50
G-Volume	12.79	94.76	3.61

H.5 CHINA: COMPLETE METHODS, 2X COSTS

All methods use the same market, chronology, factor signs, trade eligibility and execution rules. Net means and fees are in basis points per day; CAGR, maximum drawdown and expected shortfall are percentages.

Table 13: China: full-period performance at 2× costs.

Method	Net bp/d	CAGR	Sharpe	MDD	$ES_{.05}$	Fee bp/d
Fixed	0.93	-1.21	0.09	63.73	4.42	2.46
Mean	2.42	2.46	0.23	59.88	4.46	2.48
Rolling	2.07	1.67	0.20	58.88	4.42	2.23
KT	1.76	0.94	0.17	61.66	4.36	2.13
UP	2.25	2.26	0.22	55.76	4.33	2.19
UP-ratio	2.12	1.66	0.20	61.10	4.51	2.47
F-Fixed	4.03	7.05	0.39	49.94	4.19	2.02
F-Mean	2.90	3.96	0.28	57.91	4.27	2.03
F-Rolling	2.64	3.62	0.27	58.14	4.14	1.80
F-KT	3.08	4.53	0.30	56.15	4.28	1.95
F-UP	2.97	4.07	0.28	57.34	4.35	2.20
F-UP-ratio	2.98	4.18	0.29	59.44	4.36	2.03
Within	1.19	-0.59	0.11	62.36	4.43	2.43
Between	2.11	1.83	0.20	58.00	4.35	2.44
F-Rebalanced	3.09	4.86	0.32	56.33	4.12	1.43
G-Dispersion	2.46	1.67	0.21	59.57	5.05	2.07
G-Geometry	2.02	2.45	0.22	46.93	3.92	2.73
G-Position	1.35	-0.01	0.13	59.82	4.28	2.28
G-Comovement	4.16	7.97	0.44	43.93	3.91	1.57
G-Trend	1.73	0.88	0.17	58.26	4.36	2.62
G-Volume	-1.10	-5.83	-0.11	64.50	4.39	2.56

Gross traded notional, stock exposure and affected missing-mark weight are daily means in percent of pretrade NAV. Affected weight is the larger of missing-mark exposure before trading and at the label mark.

Table 14: China: turnover, exposure, and missing-mark weight at 2× costs (percent).

Method	Turnover	Exposure	Affected weight
Fixed	12.30	94.79	3.97
Mean	12.41	94.78	3.86
Rolling	11.14	94.71	4.70
KT	10.64	94.82	5.05
UP	10.96	94.84	4.70
UP-ratio	12.32	94.75	3.93
F-Fixed	10.07	94.87	5.56
F-Mean	10.16	94.84	5.67
F-Rolling	8.99	94.90	6.44
F-KT	9.73	94.89	5.95
F-UP	10.97	94.88	4.87
F-UP-ratio	10.16	94.90	5.36
Within	12.13	94.79	4.06
Between	12.17	94.81	4.03
F-Rebalanced	7.13	94.92	7.20
G-Dispersion	10.34	94.84	5.41
G-Geometry	13.65	94.87	3.37
G-Position	11.40	94.80	4.67
G-Comovement	7.86	94.90	6.63
G-Trend	13.07	94.76	3.51
G-Volume	12.78	94.75	3.62

H.6 US: COMPLETE METHODS, 1X COSTS

All methods use the same market, chronology, factor signs, trade eligibility and execution rules. Net means and fees are in basis points per day; CAGR, maximum drawdown and expected shortfall are percentages.

Table 15: US: full-period performance at $1\times$ costs.

Method	Net bp/d	CAGR	Sharpe	MDD	$ES_{.05}$	Fee bp/d
Fixed	2.12	2.36	0.22	51.95	3.56	1.85
Mean	2.93	4.38	0.30	51.97	3.57	1.84
Rolling	2.56	3.95	0.29	49.81	3.27	1.84
KT	2.90	4.73	0.32	49.19	3.34	1.83
UP	2.48	3.46	0.26	51.87	3.44	1.84
UP-ratio	2.67	3.72	0.27	52.60	3.60	1.84
F-Fixed	2.52	3.88	0.28	47.68	3.27	1.84
F-Mean	2.73	4.31	0.30	49.98	3.33	1.84
F-Rolling	2.46	3.85	0.28	48.28	3.16	1.92
F-KT	3.35	6.17	0.38	47.29	3.17	1.94
F-UP	2.66	4.31	0.30	49.36	3.21	1.84
F-UP-ratio	2.72	4.26	0.30	49.58	3.34	1.82
Within	2.35	3.03	0.25	52.40	3.48	1.84
Between	2.80	4.20	0.29	51.76	3.53	1.84
F-Rebalanced	2.67	4.43	0.31	49.31	3.18	1.84
G-Dispersion	2.17	4.16	0.33	35.76	2.55	1.74
G-Geometry	2.68	5.76	0.45	30.89	2.35	1.16
G-Position	2.78	4.19	0.29	48.76	3.46	1.85
G-Comovement	2.15	3.37	0.26	42.45	3.08	1.87
G-Trend	3.22	4.80	0.31	53.61	3.74	1.85
G-Volume	2.01	2.86	0.24	38.82	3.29	1.64

Gross traded notional, stock exposure and affected missing-mark weight are daily means in percent of pretrade NAV. Affected weight is the larger of missing-mark exposure before trading and at the label mark.

Table 16: US: turnover, exposure, and missing-mark weight at $1\times$ costs (percent).

Method	Turnover	Exposure	Affected weight
Fixed	18.46	94.99	0.55
Mean	18.41	94.99	0.60
Rolling	18.41	94.99	0.54
KT	18.27	94.99	0.60
UP	18.42	94.99	0.55
UP-ratio	18.44	94.99	0.56
F-Fixed	18.40	94.99	0.53
F-Mean	18.38	94.99	0.55
F-Rolling	19.23	94.98	0.09
F-KT	19.38	94.98	0.03
F-UP	18.37	94.99	0.55
F-UP-ratio	18.23	94.99	0.62
Within	18.44	94.99	0.56
Between	18.44	94.99	0.57
F-Rebalanced	18.34	94.99	0.54
G-Dispersion	17.38	94.99	0.80
G-Geometry	11.56	95.08	5.54
G-Position	18.46	94.99	0.59
G-Comovement	18.69	94.97	0.33
G-Trend	18.48	94.99	0.63
G-Volume	16.35	94.99	1.42

H.7 US: COMPLETE METHODS, 2X COSTS

All methods use the same market, chronology, factor signs, trade eligibility and execution rules. Net means and fees are in basis points per day; CAGR, maximum drawdown and expected shortfall are percentages.

Table 17: US: full-period performance at 2× costs.

Method	Net bp/d	CAGR	Sharpe	MDD	$ES_{.05}$	Fee bp/d
Fixed	0.27	-2.30	0.03	53.31	3.58	3.69
Mean	1.08	-0.35	0.11	52.25	3.58	3.68
Rolling	0.71	-0.77	0.08	50.12	3.29	3.68
KT	1.07	0.01	0.12	49.49	3.35	3.65
UP	0.63	-1.23	0.07	52.15	3.46	3.69
UP-ratio	0.82	-0.99	0.08	52.88	3.62	3.69
F-Fixed	0.68	-0.83	0.08	49.90	3.29	3.68
F-Mean	0.88	-0.42	0.10	50.28	3.34	3.68
F-Rolling	0.53	-1.07	0.06	50.34	3.18	3.85
F-KT	1.41	1.11	0.16	48.39	3.19	3.88
F-UP	0.82	-0.41	0.09	49.95	3.23	3.67
F-UP-ratio	0.90	-0.43	0.10	51.14	3.36	3.65
Within	0.50	-1.65	0.05	52.68	3.50	3.69
Between	0.95	-0.53	0.10	52.04	3.55	3.69
F-Rebalanced	0.83	-0.29	0.10	49.62	3.20	3.67
G-Dispersion	0.43	-0.30	0.07	35.91	2.57	3.48
G-Geometry	1.52	2.73	0.25	30.86	2.36	2.31
G-Position	0.93	-0.55	0.10	49.07	3.47	3.69
G-Comovement	0.28	-1.39	0.03	44.95	3.10	3.74
G-Trend	1.37	0.03	0.13	53.89	3.75	3.70
G-Volume	0.37	-1.29	0.04	38.94	3.30	3.27

Gross traded notional, stock exposure and affected missing-mark weight are daily means in percent of pretrade NAV. Affected weight is the larger of missing-mark exposure before trading and at the label mark.

Table 18: US: turnover, exposure, and missing-mark weight at 2× costs (percent).

Method	Turnover	Exposure	Affected weight
Fixed	18.45	94.99	0.56
Mean	18.39	94.98	0.62
Rolling	18.40	94.98	0.56
KT	18.25	94.98	0.61
UP	18.41	94.99	0.57
UP-ratio	18.43	94.98	0.57
F-Fixed	18.39	94.99	0.54
F-Mean	18.37	94.98	0.57
F-Rolling	19.22	94.98	0.09
F-KT	19.37	94.98	0.03
F-UP	18.36	94.99	0.56
F-UP-ratio	18.21	94.98	0.64
Within	18.43	94.99	0.57
Between	18.43	94.98	0.59
F-Rebalanced	18.33	94.98	0.55
G-Dispersion	17.36	94.98	0.82
G-Geometry	11.54	95.08	5.57
G-Position	18.45	94.98	0.60
G-Comovement	18.67	94.96	0.34
G-Trend	18.46	94.99	0.65
G-Volume	16.34	94.99	1.45

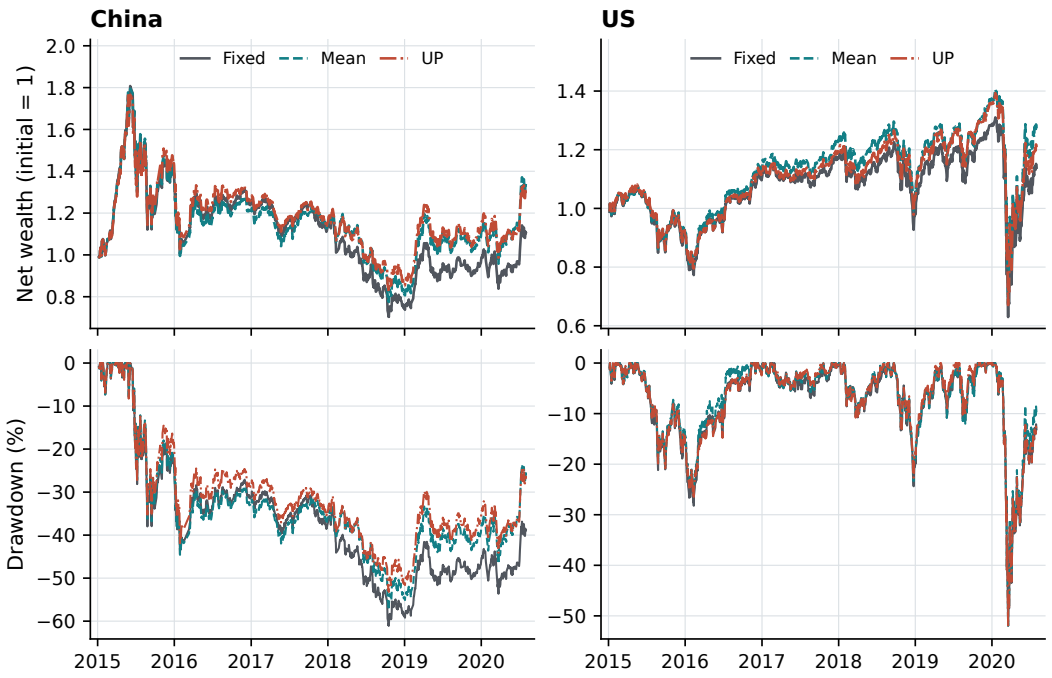

Figure 5: Net wealth and drawdown of the continuing public-market portfolios at base costs. Wealth is rebased to one at the test boundary; holdings and estimator states continue from the common 2014 warmup.

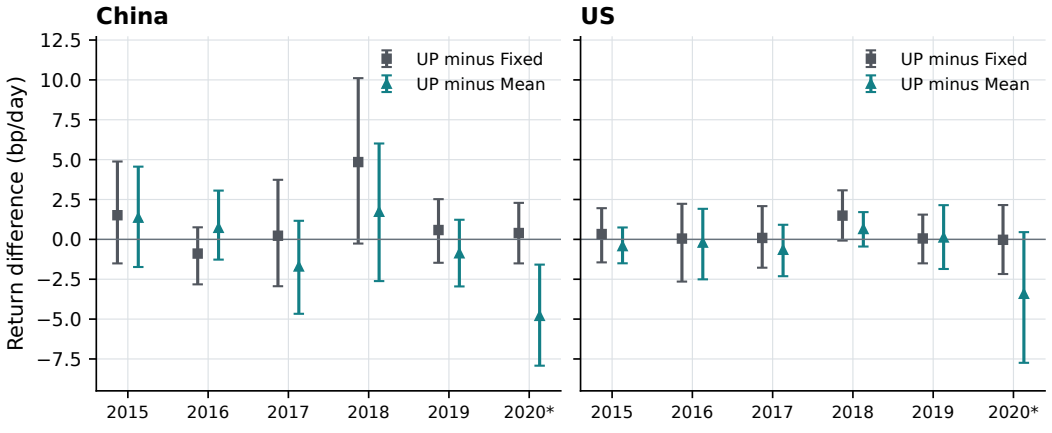

Figure 6: Annual paired UP gains and 95% block intervals against both controls. The final block, marked by an asterisk, ends in July 2020. Calendar blocks are descriptive slices of the continuing portfolio rather than independent trials.

H.8 CHINA: ANNUAL PATHS, 1X COSTS

Annualized net geometric returns (percent). The 2020 column ends on 31 July and is annualized over the observed sessions; it is not a full-year realized return.

Table 19: China: annualized geometric returns at 1× costs (percent).

Method	2015	2016	2017	2018	2019	2020
Fixed	42.52	-10.81	-9.76	-34.47	30.89	25.81
Mean	43.10	-14.41	-5.25	-29.23	35.62	42.83
Rolling	50.10	-15.07	-4.17	-32.20	33.51	31.07
KT	48.83	-14.23	-9.69	-32.18	36.57	28.69
UP	48.24	-12.89	-9.13	-25.62	32.89	27.18
UP-ratio	46.46	-13.49	-9.00	-30.99	34.93	41.70
F-Fixed	62.42	-12.17	-5.11	-25.87	38.25	42.13
F-Mean	54.76	-11.82	-7.22	-31.95	38.49	39.62
F-Rolling	46.55	-16.81	-6.52	-26.03	38.14	35.47
F-KT	47.66	-15.06	-1.11	-26.92	36.37	35.62
F-UP	55.71	-16.25	-2.62	-27.80	37.37	30.21
F-UP-ratio	45.08	-14.11	-1.93	-30.46	36.30	48.37
Within	44.47	-13.88	-9.61	-32.64	35.82	22.65
Between	46.75	-9.08	-5.55	-30.81	32.62	26.97
F-Rebalanced	47.18	-15.74	-2.54	-27.68	41.24	31.21
G-Dispersion	18.56	-21.79	7.94	-33.62	42.43	62.23
G-Geometry	43.59	-7.20	8.40	-25.77	22.72	11.28
G-Position	40.39	-13.64	-9.41	-26.88	30.84	22.51
G-Comovement	28.47	-8.24	14.70	-24.21	38.92	36.65
G-Trend	50.81	-15.13	-7.04	-26.92	32.20	19.30
G-Volume	16.51	-18.97	-0.59	-32.51	27.56	10.10

Daily arithmetic net means (bp/day) over the same calendar blocks. Each block is a slice of the continuing portfolio, with no reset of holdings or calibration.

Table 20: China: daily net-return means by year at 1× costs (bp/day).

Method	2015	2016	2017	2018	2019	2020
Fixed	17.54	-3.19	-3.74	-15.59	11.53	10.50
Mean	17.70	-4.80	-1.81	-12.48	12.99	15.69
Rolling	19.80	-5.11	-1.38	-14.37	12.29	12.13
KT	19.31	-4.71	-3.72	-14.40	13.22	11.41
UP	19.06	-4.08	-3.51	-10.75	12.12	10.89
UP-ratio	18.81	-4.38	-3.41	-13.52	12.76	15.36
F-Fixed	22.41	-3.91	-1.76	-10.77	13.74	15.30
F-Mean	20.66	-3.69	-2.66	-14.28	13.83	14.73
F-Rolling	18.01	-6.08	-2.34	-10.95	13.61	13.39
F-KT	18.71	-5.29	-0.15	-11.32	13.15	13.45
F-UP	21.24	-5.96	-0.73	-11.75	13.43	11.77
F-UP-ratio	18.15	-4.86	-0.49	-13.29	13.17	17.15
Within	18.16	-4.55	-3.68	-14.49	12.99	9.44
Between	18.61	-2.43	-1.92	-13.45	12.05	10.86
F-Rebalanced	18.41	-5.68	-0.80	-11.94	14.48	12.07
G-Dispersion	11.25	-8.12	3.42	-14.93	15.20	21.30
G-Geometry	17.34	-2.04	3.35	-11.00	8.67	5.29
G-Position	16.79	-4.44	-3.60	-11.36	11.48	9.42
G-Comovement	12.83	-2.54	5.66	-10.05	13.69	13.64
G-Trend	19.89	-5.15	-2.58	-11.36	11.83	8.31
G-Volume	9.66	-7.22	-0.01	-14.95	10.54	5.20

H.9 CHINA: ANNUAL PATHS, 2X COSTS

Annualized net geometric returns (percent). The 2020 column ends on 31 July and is annualized over the observed sessions; it is not a full-year realized return.

Table 21: China: annualized geometric returns at 2× costs (percent).

Method	2015	2016	2017	2018	2019	2020
Fixed	38.19	-13.37	-10.59	-36.52	25.07	20.18
Mean	38.70	-16.82	-6.23	-31.51	29.65	36.49
Rolling	44.85	-17.54	-4.72	-33.21	27.58	25.26
KT	44.30	-16.65	-10.20	-33.10	30.50	22.96
UP	43.61	-15.56	-9.75	-26.63	27.02	21.49
UP-ratio	41.45	-15.91	-9.89	-33.00	29.03	35.38
F-Fixed	58.83	-13.76	-5.48	-27.71	32.11	35.83
F-Mean	50.16	-13.78	-7.60	-32.97	32.37	33.44
F-Rolling	44.02	-18.87	-6.96	-26.91	32.55	29.49
F-KT	44.82	-16.44	-1.52	-28.66	30.16	29.65
F-UP	53.10	-17.80	-3.71	-30.27	31.23	24.46
F-UP-ratio	42.41	-15.77	-2.68	-32.14	30.22	41.79
Within	39.98	-16.37	-10.43	-34.60	29.88	17.14
Between	42.31	-11.69	-6.34	-32.91	26.62	21.29
F-Rebalanced	44.97	-16.20	-2.67	-28.45	35.58	25.49
G-Dispersion	16.90	-23.14	7.06	-35.89	36.23	55.41
G-Geometry	40.79	-10.71	6.17	-28.54	17.14	5.99
G-Position	37.24	-16.06	-10.52	-28.69	25.00	17.08
G-Comovement	26.85	-9.69	14.35	-24.94	33.43	30.31
G-Trend	46.18	-17.73	-8.84	-29.09	26.31	13.97
G-Volume	12.53	-21.76	-3.39	-33.31	21.99	5.34

Daily arithmetic net means (bp/day) over the same calendar blocks. Each block is a slice of the continuing portfolio, with no reset of holdings or calibration.

Table 22: China: daily net-return means by year at 2× costs (bp/day).

Method	2015	2016	2017	2018	2019	2020
Fixed	16.32	-4.34	-4.10	-16.85	9.73	8.68
Mean	16.46	-5.94	-2.22	-13.78	11.21	13.89
Rolling	18.38	-6.28	-1.61	-14.96	10.48	10.33
KT	18.08	-5.85	-3.94	-14.94	11.42	9.60
UP	17.79	-5.32	-3.78	-11.29	10.32	9.08
UP-ratio	17.43	-5.51	-3.79	-14.69	10.98	13.55
F-Fixed	21.52	-4.64	-1.92	-11.76	11.94	13.50
F-Mean	19.46	-4.58	-2.83	-14.88	12.03	12.93
F-Rolling	17.32	-7.07	-2.52	-11.43	11.97	11.60
F-KT	17.94	-5.94	-0.32	-12.27	11.30	11.67
F-UP	20.57	-6.70	-1.17	-13.13	11.62	9.98
F-UP-ratio	17.40	-5.64	-0.80	-14.26	11.35	15.34
Within	16.91	-5.71	-4.04	-15.66	11.21	7.61
Between	17.39	-3.58	-2.25	-14.67	10.21	9.04
F-Rebalanced	17.81	-5.90	-0.86	-12.36	12.86	10.30
G-Dispersion	10.69	-8.81	3.09	-16.31	13.43	19.59
G-Geometry	16.56	-3.56	2.52	-12.51	6.82	3.36
G-Position	15.89	-5.57	-4.09	-12.35	9.67	7.62
G-Comovement	12.33	-3.17	5.54	-10.43	12.09	11.75
G-Trend	18.65	-6.39	-3.36	-12.55	10.02	6.49
G-Volume	8.28	-8.61	-1.15	-15.42	8.77	3.44

H.10 US: ANNUAL PATHS, 1X COSTS

Annualized net geometric returns (percent). The 2020 column ends on 31 July and is annualized over the observed sessions; it is not a full-year realized return.

Table 23: US: annualized geometric returns at 1× costs (percent).

Method	2015	2016	2017	2018	2019	2020
Fixed	-9.75	21.33	6.80	-16.48	31.58	-18.90
Mean	-7.89	22.24	8.88	-14.62	31.32	-12.82
Rolling	-7.94	18.79	8.63	-12.92	29.18	-12.12
KT	-6.76	18.73	9.01	-12.91	29.81	-8.87
UP	-8.87	21.84	7.15	-13.22	31.81	-18.54
UP-ratio	-8.05	21.63	8.24	-15.05	35.82	-20.25
F-Fixed	-6.28	14.34	7.69	-16.25	30.36	-3.38
F-Mean	-5.91	19.05	10.89	-13.79	27.67	-12.60
F-Rolling	-5.30	14.68	9.76	-15.40	26.14	-5.13
F-KT	-2.06	14.89	12.24	-14.58	27.52	2.46
F-UP	-4.80	14.39	14.94	-14.02	26.21	-11.54
F-UP-ratio	-7.29	19.08	11.57	-15.80	26.44	-6.57
Within	-9.56	20.01	7.63	-13.94	32.36	-18.69
Between	-8.37	23.55	10.74	-13.82	33.11	-20.70
F-Rebalanced	-4.90	14.78	15.81	-12.53	27.80	-16.48
G-Dispersion	-1.73	7.55	12.96	-4.27	17.86	-11.41
G-Geometry	-0.27	6.48	12.93	-5.45	25.51	-6.73
G-Position	-8.24	21.02	4.28	-16.54	36.54	-8.24
G-Comovement	-6.77	11.25	11.19	-11.11	23.98	-9.12
G-Trend	-12.37	23.34	7.68	-13.34	40.68	-14.31
G-Volume	-2.81	10.38	6.97	-8.59	24.96	-17.75

Daily arithmetic net means (bp/day) over the same calendar blocks. Each block is a slice of the continuing portfolio, with no reset of holdings or calibration.

Table 24: US: daily net-return means by year at 1× costs (bp/day).

Method	2015	2016	2017	2018	2019	2020
Fixed	-3.44	8.26	2.79	-6.59	11.31	-0.99
Mean	-2.68	8.51	3.53	-5.75	11.25	2.39
Rolling	-2.73	7.26	3.42	-4.98	10.52	1.22
KT	-2.23	7.25	3.55	-4.98	10.74	3.09
UP	-3.10	8.31	2.88	-5.10	11.37	-1.02
UP-ratio	-2.73	8.35	3.28	-5.94	12.59	-1.25
F-Fixed	-2.01	5.81	3.10	-6.52	10.88	4.70
F-Mean	-1.90	7.38	4.24	-5.39	10.07	1.30
F-Rolling	-1.66	5.81	3.82	-6.13	9.56	3.88
F-KT	-0.32	5.90	4.71	-5.77	10.01	7.07
F-UP	-1.44	5.74	5.66	-5.49	9.59	1.23
F-UP-ratio	-2.49	7.40	4.48	-6.32	9.69	4.08
Within	-3.40	7.74	3.07	-5.42	11.54	-0.87
Between	-2.85	8.95	4.22	-5.36	11.77	-1.98
F-Rebalanced	-1.49	5.86	5.95	-4.84	10.09	-1.22
G-Dispersion	-0.26	3.16	4.92	-1.39	6.72	-1.80
G-Geometry	0.24	2.72	4.89	-1.95	9.17	-0.27
G-Position	-2.76	8.19	1.86	-6.58	12.79	3.31
G-Comovement	-2.25	4.69	4.36	-4.16	8.87	0.77
G-Trend	-4.55	8.98	3.15	-5.08	14.03	2.28
G-Volume	-0.67	4.35	2.77	-3.13	9.15	-2.20

H.11 US: ANNUAL PATHS, 2X COSTS

Annualized net geometric returns (percent). The 2020 column ends on 31 July and is annualized over the observed sessions; it is not a full-year realized return.

Table 25: US: annualized geometric returns at 2× costs (percent).

Method	2015	2016	2017	2018	2019	2020
Fixed	-14.08	15.53	1.69	-20.48	26.37	-22.09
Mean	-12.24	16.40	3.68	-18.70	26.11	-16.26
Rolling	-12.33	13.12	3.47	-17.08	24.06	-15.60
KT	-11.21	13.07	3.83	-17.07	24.67	-12.17
UP	-13.21	16.02	2.05	-17.37	26.59	-21.74
UP-ratio	-12.43	15.83	3.08	-19.11	30.43	-23.39
F-Fixed	-10.75	8.90	2.55	-20.26	25.20	-7.19
F-Mean	-10.35	13.37	5.60	-17.92	22.62	-16.04
F-Rolling	-9.80	9.24	4.54	-19.44	20.14	-9.48
F-KT	-6.71	9.43	6.91	-18.66	21.45	-2.53
F-UP	-9.32	8.96	9.48	-18.13	21.23	-15.03
F-UP-ratio	-11.68	13.41	6.27	-19.82	21.45	-9.96
Within	-13.87	14.28	2.49	-18.06	27.12	-21.90
Between	-12.75	17.66	5.46	-17.95	27.84	-23.80
F-Rebalanced	-9.39	9.34	10.32	-16.70	22.75	-19.77
G-Dispersion	-6.31	2.52	7.66	-8.73	13.47	-13.66
G-Geometry	-4.87	1.52	7.63	-7.28	25.48	-6.72
G-Position	-12.62	15.23	-0.70	-20.53	31.10	-11.88
G-Comovement	-11.11	6.00	5.95	-15.31	18.21	-12.59
G-Trend	-16.52	17.43	2.52	-17.50	35.07	-17.70
G-Volume	-7.37	5.22	1.97	-12.22	21.38	-20.07

Daily arithmetic net means (bp/day) over the same calendar blocks. Each block is a slice of the continuing portfolio, with no reset of holdings or calibration.

Table 26: US: daily net-return means by year at 2× costs (bp/day).

Method	2015	2016	2017	2018	2019	2020
Fixed	-5.39	6.32	0.84	-8.53	9.71	-2.60
Mean	-4.60	6.57	1.59	-7.69	9.64	0.78
Rolling	-4.66	5.32	1.48	-6.92	8.92	-0.39
KT	-4.17	5.31	1.62	-6.92	9.14	1.61
UP	-5.04	6.37	0.94	-7.04	9.76	-2.62
UP-ratio	-4.66	6.41	1.34	-7.88	10.98	-2.86
F-Fixed	-3.95	3.88	1.16	-8.46	9.28	3.10
F-Mean	-3.83	5.44	2.30	-7.33	8.47	-0.31
F-Rolling	-3.59	3.88	1.89	-8.07	7.62	2.02
F-KT	-2.25	3.97	2.78	-7.71	8.07	5.10
F-UP	-3.36	3.81	3.73	-7.43	8.00	-0.38
F-UP-ratio	-4.42	5.46	2.54	-8.26	8.09	2.61
Within	-5.34	5.80	1.13	-7.37	9.93	-2.48
Between	-4.79	7.01	2.27	-7.31	10.16	-3.58
F-Rebalanced	-3.41	3.93	4.01	-6.78	8.49	-2.83
G-Dispersion	-2.16	1.27	3.01	-3.29	5.22	-2.83
G-Geometry	-1.63	0.83	2.98	-2.73	9.16	-0.27
G-Position	-4.70	6.24	-0.08	-8.53	11.18	1.69
G-Comovement	-4.14	2.77	2.44	-6.08	6.98	-0.78
G-Trend	-6.47	7.03	1.20	-7.03	12.41	0.66
G-Volume	-2.57	2.45	0.86	-4.74	8.00	-3.36

H.12 CHINA: PAIRED COMPARISONS, 1X COSTS

Differences are paired daily net-return means (bp/day). Intervals use 2,000 circular moving-block resamples of length 20. Holm diagnostics compare five adaptive methods with Fixed or four interval methods with Mean within each architecture, market and cost setting; they do not adjust across all markets and years.

Table 27: China: full-period paired return differences at $1\times$ costs (bp/day).

Method	Comparator	Δ bp/d	95% interval	Holm p
Mean	Fixed	+1.50	[0.08, 3.17]	0.190
Rolling	Fixed	+1.03	[-0.48, 2.63]	0.368
KT	Fixed	+0.67	[-0.79, 2.08]	0.370
UP	Fixed	+1.19	[-0.27, 2.72]	0.294
UP-ratio	Fixed	+1.19	[0.05, 2.50]	0.190
Rolling	Mean	-0.47	[-2.25, 1.30]	1.000
KT	Mean	-0.82	[-2.45, 0.64]	1.000
UP	Mean	-0.31	[-1.67, 1.09]	1.000
UP-ratio	Mean	-0.30	[-1.71, 1.05]	1.000
F-Mean	F-Fixed	-1.12	[-2.51, 0.20]	0.368
F-Rolling	F-Fixed	-1.49	[-3.10, -0.05]	0.225
F-KT	F-Fixed	-0.98	[-2.64, 0.37]	0.552
F-UP	F-Fixed	-0.97	[-2.91, 0.94]	0.552
F-UP-ratio	F-Fixed	-1.04	[-3.03, 0.61]	0.552
F-Rolling	F-Mean	-0.37	[-2.22, 1.25]	1.000
F-KT	F-Mean	+0.14	[-1.82, 1.84]	1.000
F-UP	F-Mean	+0.15	[-2.34, 2.44]	1.000
F-UP-ratio	F-Mean	+0.08	[-2.13, 1.83]	1.000

The primary UP comparisons by calendar block are shown below. The 2020 block is January–July. The supplementary CSV also retains all annual comparator variants.

Table 28: China: paired UP return differences by year at $1\times$ costs (bp/day).

Year	Method	Comparator	Δ bp/d	95% interval
2015	UP	Fixed	+1.51	[-1.51, 4.88]
2015	UP	Mean	+1.36	[-1.73, 4.56]
2015	F-UP	F-Fixed	-1.17	[-10.08, 7.09]
2015	F-UP	F-Mean	+0.58	[-10.15, 10.35]
2016	UP	Fixed	-0.89	[-2.82, 0.76]
2016	UP	Mean	+0.72	[-1.27, 3.06]
2016	F-UP	F-Fixed	-2.05	[-5.91, 1.12]
2016	F-UP	F-Mean	-2.26	[-6.88, 1.68]
2017	UP	Fixed	+0.23	[-2.94, 3.73]
2017	UP	Mean	-1.70	[-4.67, 1.16]
2017	F-UP	F-Fixed	+1.03	[-2.40, 4.65]
2017	F-UP	F-Mean	+1.93	[-1.38, 5.30]
2018	UP	Fixed	+4.85	[-0.26, 10.11]
2018	UP	Mean	+1.73	[-2.62, 6.01]
2018	F-UP	F-Fixed	-0.99	[-4.25, 2.65]
2018	F-UP	F-Mean	+2.53	[-2.61, 7.97]
2019	UP	Fixed	+0.59	[-1.47, 2.52]
2019	UP	Mean	-0.87	[-2.95, 1.23]
2019	F-UP	F-Fixed	-0.31	[-1.70, 1.25]
2019	F-UP	F-Mean	-0.39	[-2.57, 1.73]
2020	UP	Fixed	+0.40	[-1.51, 2.29]
2020	UP	Mean	-4.80	[-7.92, -1.58]
2020	F-UP	F-Fixed	-3.53	[-5.36, -1.81]
2020	F-UP	F-Mean	-2.96	[-6.23, 0.65]

H.13 CHINA: PAIRED COMPARISONS, 2X COSTS

Differences are paired daily net-return means (bp/day). Intervals use 2,000 circular moving-block resamples of length 20. Holm diagnostics compare five adaptive methods with Fixed or four interval methods with Mean within each architecture, market and cost setting; they do not adjust across all markets and years.

Table 29: China: full-period paired return differences at 2× costs (bp/day).

Method	Comparator	Δ bp/d	95% interval	Holm p
Mean	Fixed	+1.49	[0.07, 3.16]	0.190
Rolling	Fixed	+1.14	[-0.37, 2.75]	0.280
KT	Fixed	+0.83	[-0.64, 2.26]	0.280
UP	Fixed	+1.32	[-0.17, 2.89]	0.225
UP-ratio	Fixed	+1.19	[0.06, 2.48]	0.190
Rolling	Mean	-0.34	[-2.10, 1.39]	1.000
KT	Mean	-0.65	[-2.26, 0.81]	1.000
UP	Mean	-0.17	[-1.55, 1.25]	1.000
UP-ratio	Mean	-0.29	[-1.71, 1.06]	1.000
F-Mean	F-Fixed	-1.13	[-2.55, 0.21]	0.368
F-Rolling	F-Fixed	-1.38	[-2.97, 0.05]	0.290
F-KT	F-Fixed	-0.95	[-2.61, 0.40]	0.597
F-UP	F-Fixed	-1.06	[-3.01, 0.83]	0.597
F-UP-ratio	F-Fixed	-1.04	[-3.02, 0.61]	0.597
F-Rolling	F-Mean	-0.25	[-2.10, 1.37]	1.000
F-KT	F-Mean	+0.18	[-1.78, 1.90]	1.000
F-UP	F-Mean	+0.07	[-2.44, 2.42]	1.000
F-UP-ratio	F-Mean	+0.09	[-2.13, 1.85]	1.000

The primary UP comparisons by calendar block are shown below. The 2020 block is January–July. The supplementary CSV also retains all annual comparator variants.

Table 30: China: paired UP return differences by year at 2× costs (bp/day).

Year	Method	Comparator	Δ bp/d	95% interval
2015	UP	Fixed	+1.48	[-1.51, 4.81]
2015	UP	Mean	+1.34	[-1.75, 4.55]
2015	F-UP	F-Fixed	-0.95	[-9.92, 7.30]
2015	F-UP	F-Mean	+1.11	[-9.66, 10.99]
2016	UP	Fixed	-0.97	[-2.98, 0.71]
2016	UP	Mean	+0.62	[-1.39, 2.97]
2016	F-UP	F-Fixed	-2.06	[-5.86, 1.18]
2016	F-UP	F-Mean	-2.12	[-6.54, 1.83]
2017	UP	Fixed	+0.32	[-2.82, 3.80]
2017	UP	Mean	-1.56	[-4.56, 1.38]
2017	F-UP	F-Fixed	+0.74	[-2.70, 4.42]
2017	F-UP	F-Mean	+1.65	[-1.69, 5.02]
2018	UP	Fixed	+5.56	[0.39, 10.86]
2018	UP	Mean	+2.49	[-1.96, 6.78]
2018	F-UP	F-Fixed	-1.37	[-4.62, 2.31]
2018	F-UP	F-Mean	+1.75	[-3.34, 7.23]
2019	UP	Fixed	+0.60	[-1.51, 2.56]
2019	UP	Mean	-0.88	[-2.98, 1.23]
2019	F-UP	F-Fixed	-0.32	[-1.71, 1.23]
2019	F-UP	F-Mean	-0.41	[-2.59, 1.71]
2020	UP	Fixed	+0.40	[-1.50, 2.30]
2020	UP	Mean	-4.81	[-7.95, -1.59]
2020	F-UP	F-Fixed	-3.52	[-5.36, -1.81]
2020	F-UP	F-Mean	-2.95	[-6.23, 0.66]

H.14 US: PAIRED COMPARISONS, 1X COSTS

Differences are paired daily net-return means (bp/day). Intervals use 2,000 circular moving-block resamples of length 20. Holm diagnostics compare five adaptive methods with Fixed or four interval methods with Mean within each architecture, market and cost setting; they do not adjust across all markets and years.

Table 31: US: full-period paired return differences at $1\times$ costs (bp/day).

Method	Comparator	Δ bp/d	95% interval	Holm p
Mean	Fixed	+0.81	[0.10, 1.54]	0.125
Rolling	Fixed	+0.44	[-0.77, 1.60]	0.826
KT	Fixed	+0.78	[-0.38, 2.11]	0.744
UP	Fixed	+0.36	[-0.48, 1.11]	0.826
UP-ratio	Fixed	+0.55	[-0.25, 1.29]	0.744
Rolling	Mean	-0.37	[-1.60, 0.77]	1.000
KT	Mean	-0.02	[-1.17, 1.14]	1.000
UP	Mean	-0.45	[-1.35, 0.39]	1.000
UP-ratio	Mean	-0.26	[-1.01, 0.38]	1.000
F-Mean	F-Fixed	+0.21	[-0.65, 1.04]	1.000
F-Rolling	F-Fixed	-0.06	[-1.12, 1.00]	1.000
F-KT	F-Fixed	+0.83	[-0.22, 1.78]	0.600
F-UP	F-Fixed	+0.14	[-0.89, 1.13]	1.000
F-UP-ratio	F-Fixed	+0.20	[-0.63, 1.09]	1.000
F-Rolling	F-Mean	-0.27	[-1.26, 0.78]	1.000
F-KT	F-Mean	+0.62	[-0.44, 1.75]	0.932
F-UP	F-Mean	-0.07	[-0.85, 0.73]	1.000
F-UP-ratio	F-Mean	-0.00	[-0.53, 0.54]	1.000

The primary UP comparisons by calendar block are shown below. The 2020 block is January–July. The supplementary CSV also retains all annual comparator variants.

Table 32: US: paired UP return differences by year at $1\times$ costs (bp/day).

Year	Method	Comparator	Δ bp/d	95% interval
2015	UP	Fixed	+0.33	[-1.44, 1.96]
2015	UP	Mean	-0.42	[-1.50, 0.75]
2015	F-UP	F-Fixed	+0.58	[-1.69, 2.98]
2015	F-UP	F-Mean	+0.47	[-0.79, 1.86]
2016	UP	Fixed	+0.05	[-2.65, 2.23]
2016	UP	Mean	-0.20	[-2.51, 1.92]
2016	F-UP	F-Fixed	-0.07	[-2.60, 2.57]
2016	F-UP	F-Mean	-1.64	[-3.99, 0.25]
2017	UP	Fixed	+0.09	[-1.78, 2.09]
2017	UP	Mean	-0.65	[-2.31, 0.91]
2017	F-UP	F-Fixed	+2.56	[0.63, 4.41]
2017	F-UP	F-Mean	+1.42	[0.06, 2.75]
2018	UP	Fixed	+1.49	[-0.08, 3.08]
2018	UP	Mean	+0.65	[-0.45, 1.71]
2018	F-UP	F-Fixed	+1.03	[-0.46, 2.41]
2018	F-UP	F-Mean	-0.10	[-1.44, 1.27]
2019	UP	Fixed	+0.06	[-1.50, 1.55]
2019	UP	Mean	+0.12	[-1.86, 2.14]
2019	F-UP	F-Fixed	-1.29	[-2.84, 0.15]
2019	F-UP	F-Mean	-0.48	[-1.83, 0.92]
2020	UP	Fixed	-0.03	[-2.17, 2.15]
2020	UP	Mean	-3.41	[-7.74, 0.45]
2020	F-UP	F-Fixed	-3.48	[-8.23, 0.25]
2020	F-UP	F-Mean	-0.07	[-4.24, 3.26]

H.15 US: PAIRED COMPARISONS, 2X COSTS

Differences are paired daily net-return means (bp/day). Intervals use 2,000 circular moving-block resamples of length 20. Holm diagnostics compare five adaptive methods with Fixed or four interval methods with Mean within each architecture, market and cost setting; they do not adjust across all markets and years.

Table 33: US: full-period paired return differences at 2× costs (bp/day).

Method	Comparator	Δ bp/d	95% interval	Holm p
Mean	Fixed	+0.81	[0.10, 1.55]	0.120
Rolling	Fixed	+0.44	[-0.77, 1.60]	0.812
KT	Fixed	+0.80	[-0.35, 2.12]	0.740
UP	Fixed	+0.36	[-0.48, 1.12]	0.812
UP-ratio	Fixed	+0.55	[-0.25, 1.30]	0.740
Rolling	Mean	-0.37	[-1.61, 0.77]	1.000
KT	Mean	-0.01	[-1.16, 1.13]	1.000
UP	Mean	-0.45	[-1.35, 0.39]	1.000
UP-ratio	Mean	-0.26	[-1.01, 0.38]	1.000
F-Mean	F-Fixed	+0.21	[-0.65, 1.04]	1.000
F-Rolling	F-Fixed	-0.14	[-1.19, 0.92]	1.000
F-KT	F-Fixed	+0.74	[-0.31, 1.70]	0.845
F-UP	F-Fixed	+0.14	[-0.89, 1.14]	1.000
F-UP-ratio	F-Fixed	+0.22	[-0.61, 1.10]	1.000
F-Rolling	F-Mean	-0.35	[-1.35, 0.69]	1.000
F-KT	F-Mean	+0.53	[-0.54, 1.64]	1.000
F-UP	F-Mean	-0.07	[-0.84, 0.73]	1.000
F-UP-ratio	F-Mean	+0.01	[-0.52, 0.56]	1.000

The primary UP comparisons by calendar block are shown below. The 2020 block is January–July. The supplementary CSV also retains all annual comparator variants.

Table 34: US: paired UP return differences by year at 2× costs (bp/day).

Year	Method	Comparator	Δ bp/d	95% interval
2015	UP	Fixed	+0.35	[-1.42, 1.98]
2015	UP	Mean	-0.44	[-1.53, 0.73]
2015	F-UP	F-Fixed	+0.59	[-1.68, 2.99]
2015	F-UP	F-Mean	+0.46	[-0.80, 1.85]
2016	UP	Fixed	+0.05	[-2.65, 2.23]
2016	UP	Mean	-0.20	[-2.51, 1.92]
2016	F-UP	F-Fixed	-0.07	[-2.58, 2.57]
2016	F-UP	F-Mean	-1.63	[-3.97, 0.26]
2017	UP	Fixed	+0.09	[-1.77, 2.09]
2017	UP	Mean	-0.65	[-2.31, 0.92]
2017	F-UP	F-Fixed	+2.57	[0.64, 4.42]
2017	F-UP	F-Mean	+1.42	[0.06, 2.76]
2018	UP	Fixed	+1.49	[-0.07, 3.08]
2018	UP	Mean	+0.65	[-0.45, 1.71]
2018	F-UP	F-Fixed	+1.03	[-0.45, 2.42]
2018	F-UP	F-Mean	-0.10	[-1.44, 1.27]
2019	UP	Fixed	+0.06	[-1.50, 1.55]
2019	UP	Mean	+0.12	[-1.86, 2.14]
2019	F-UP	F-Fixed	-1.28	[-2.84, 0.15]
2019	F-UP	F-Mean	-0.48	[-1.83, 0.93]
2020	UP	Fixed	-0.02	[-2.16, 2.16]
2020	UP	Mean	-3.40	[-7.73, 0.47]
2020	F-UP	F-Fixed	-3.48	[-8.23, 0.25]
2020	F-UP	F-Mean	-0.07	[-4.23, 3.26]

H.16 ALL OBSERVED-LABEL CALIBRATION RESULTS

The same complete-label mask and mean forecast serve all three radius updates. Readiness counts 60 complete observations across lanes; a particular lane can therefore have a shorter observed history. Coverage uses each issued signed radius and its issue-time mean. The mean nonnegative width measures the set shown to an allocation rule; the raw signed widths remain in the accompanying CSV. Missing-label days are not counted as covered. Calibration rows include only complete labels with issue-time readiness, whereas financial rows retain all test sessions.

Table 35: Observed-label calibration by market and year. Width is the nonnegative set width in basis points. Negative signed radii define empty sets; Empty reports their frequency.

Market	Period	Update	Labels	Coverage %	Width bp	Empty %
China	ALL	Rolling	53,230	88.81	451.5	0.000
China	ALL	KT	53,230	81.75	354.8	0.000
China	ALL	UP	53,230	83.54	440.8	0.000
China	2015	Rolling	3,852	83.46	557.2	0.000
China	2015	KT	3,852	72.61	424.3	0.000
China	2015	UP	3,852	79.75	622.6	0.000
China	2016	Rolling	2,757	89.95	570.1	0.000
China	2016	KT	2,757	83.03	402.9	0.000
China	2016	UP	2,757	84.37	533.2	0.000
China	2017	Rolling	702	96.87	483.1	0.000
China	2017	KT	702	89.89	295.8	0.000
China	2017	UP	702	85.33	312.0	0.000
China	2018	Rolling	2,984	91.49	490.7	0.000
China	2018	KT	2,984	83.85	359.8	0.000
China	2018	UP	2,984	84.99	443.4	0.000
China	2019	Rolling	25,788	92.01	413.4	0.000
China	2019	KT	25,788	85.56	336.5	0.000
China	2019	UP	25,788	84.52	380.0	0.000
China	2020	Rolling	17,147	84.24	458.0	0.000
China	2020	KT	17,147	77.18	360.7	0.000
China	2020	UP	17,147	82.46	481.5	0.000
US	ALL	Rolling	143,895	89.35	330.5	0.000
US	ALL	KT	143,895	85.35	267.8	0.000
US	ALL	UP	143,895	86.23	313.4	0.000
US	2015	Rolling	30,379	90.83	371.8	0.000
US	2015	KT	30,379	83.16	283.7	0.000
US	2015	UP	30,379	85.04	355.8	0.000
US	2016	Rolling	33,953	90.79	359.0	0.000
US	2016	KT	33,953	89.66	324.8	0.000
US	2016	UP	33,953	87.53	322.6	0.000
US	2017	Rolling	35,973	90.96	193.4	0.000
US	2017	KT	35,973	90.14	186.1	0.000
US	2017	UP	35,973	87.16	188.0	0.000
US	2018	Rolling	27,003	84.11	278.9	0.000
US	2018	KT	27,003	77.48	229.2	0.000
US	2018	UP	27,003	83.75	293.8	0.000
US	2019	Rolling	11,113	92.33	352.0	0.000
US	2019	KT	11,113	90.32	299.5	0.000
US	2019	UP	11,113	90.65	334.1	0.000
US	2020	Rolling	5,474	81.48	1035.5	0.000
US	2020	KT	5,474	68.12	487.9	0.000
US	2020	UP	5,474	81.93	899.7	0.000

H.17 UNCERTAINTY OF THE RISK COMPARISONS

This diagnostic was added after the predefined replay to quantify uncertainty in the reported downside-risk point estimates. It uses the same 20-session circular blocks, 2,000 resamples and seed as the paired mean-return analysis. Each resample uses identical dates for the two compared return series and recomputes both risk metrics. Drawdown is path-dependent; the interval describes the resampled block paths, not a bound under arbitrary future market changes. No strategy is selected or changed using these diagnostics.

Differences below are UP minus the indicated comparator, in percentage points; negative means lower loss. The Chinese UP–Fixed drawdown improvement is not separated from zero by this interval, despite its 7.87-point estimate. US expected shortfall has a negative interval against both controls. No adjustment is made across these additional risk comparisons. The original joint efficacy criterion remains failed because of the mean-return comparisons.

Table 36: Exploratory UP risk differences by market, cost setting, and combiner. Values and 95% intervals are in percentage points (pp).

Market	Costs	Rule	Comparator	Risk	Δ pp	95% interval
China	1	All	Fixed	MDD	-7.873	[-10.442, 1.442]
China	1	All	Fixed	ES	-0.083	[-0.190, 0.027]
China	1	All	Mean	MDD	-3.640	[-5.621, 3.920]
China	1	All	Mean	ES	-0.126	[-0.224, -0.030]
China	1	Family	Fixed	MDD	+7.182	[-2.382, 12.301]
China	1	Family	Fixed	ES	+0.161	[-0.071, 0.411]
China	1	Family	Mean	MDD	-0.711	[-6.048, 10.591]
China	1	Family	Mean	ES	+0.087	[-0.175, 0.340]
China	2	All	Fixed	MDD	-7.969	[-10.716, 1.255]
China	2	All	Fixed	ES	-0.083	[-0.191, 0.026]
China	2	All	Mean	MDD	-4.115	[-6.157, 3.895]
China	2	All	Mean	ES	-0.127	[-0.226, -0.031]
China	2	Family	Fixed	MDD	+7.398	[-2.302, 12.350]
China	2	Family	Fixed	ES	+0.161	[-0.072, 0.411]
China	2	Family	Mean	MDD	-0.572	[-6.175, 10.658]
China	2	Family	Mean	ES	+0.084	[-0.179, 0.339]
US	1	All	Fixed	MDD	-0.072	[-5.346, 0.629]
US	1	All	Fixed	ES	-0.122	[-0.204, -0.052]
US	1	All	Mean	MDD	-0.092	[-3.238, 2.169]
US	1	All	Mean	ES	-0.123	[-0.216, -0.041]
US	1	Family	Fixed	MDD	+1.681	[-5.688, 3.021]
US	1	Family	Fixed	ES	-0.058	[-0.164, 0.064]
US	1	Family	Mean	MDD	-0.626	[-3.828, 1.508]
US	1	Family	Mean	ES	-0.115	[-0.202, -0.042]
US	2	All	Fixed	MDD	-1.155	[-5.870, 0.668]
US	2	All	Fixed	ES	-0.122	[-0.204, -0.052]
US	2	All	Mean	MDD	-0.091	[-3.196, 2.519]
US	2	All	Mean	ES	-0.123	[-0.215, -0.041]
US	2	Family	Fixed	MDD	+0.050	[-6.642, 2.936]
US	2	Family	Fixed	ES	-0.058	[-0.165, 0.064]
US	2	Family	Mean	MDD	-0.327	[-4.167, 1.554]
US	2	Family	Mean	ES	-0.114	[-0.201, -0.042]

H.18 NUMERICAL AND IMPLEMENTATION VERIFICATION

An independent reconstruction recovers every native UP radius to absolute error below 8×10^{-16} and log wealth to error below 1.1×10^{-12}. Every observed full-period and annual lane sequence satisfies the continuing-window bound. The certificate file records each factor, lane, window, observation count, miscoverage and bound, including wide bounds. Recomputed financial metrics agree to absolute error below 4×10^{-15}; the maximum fee-identity residual is below 7×10^{-16}. Future-return perturbations preserve all coefficients whose release histories are unchanged, and warmup portfolios match Fixed exactly.

An initial replay used a scale-dependent share-count threshold. Every displayed result comes from the corrected executor, which passes a 10^{24} price-unit rescaling test. Both markets and all methods were rerun without changing feature signs, scientific parameters, chronological boundaries or costs.

I Complete Exploratory Fractional-Feedback Replay

The new arms were specified in the dated protocol before their results were read; the prior test years and original method outcomes were already known. All contrasts below are exploratory. IUP weights issue-time ALL-score sensitivities; UUP uses uniform factor weights. Both use the same factor library, release calendar, stock-score combiner, continuing holdings and fee ledger as Fixed, Mean and original factorwise UP.

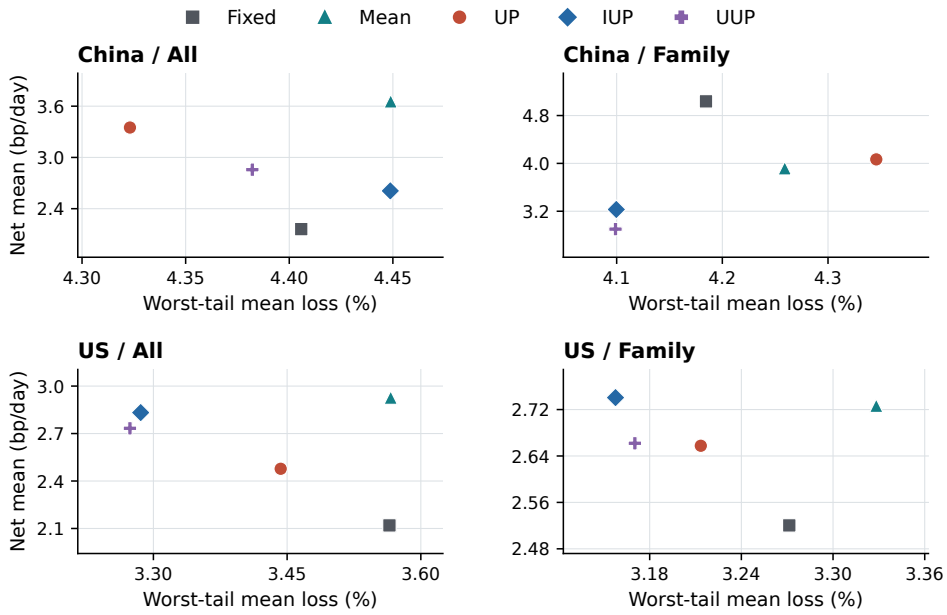

Figure 7: All five frozen arms under each market and score combiner at base costs. Vertical axis: net basis points per day; horizontal axis: worst-5% mean loss in percent. Higher net mean and lower loss are preferred. This descriptive plot includes the unfavorable Family outcome.

Observed calibration and partial identification. On released test feedback, IUP has 1,181 scored Chinese issues and 1,400 US issues. The average observed sensitivity mass is .2941 and .6420. The conditional misses are .1395 and .1294; original factorwise UP on those same scored issues and weights has .1775 and .1366. The last three test issues have complete outcomes but no release inside the update horizon; China also has 141 issues with no complete factor label. On the common released subset the full-factor miss can be anywhere in $[.0408, .7467]$ in China and $[.0826, .4406]$ in the US. These are conservative identification bounds, not 95% statistical intervals. The original UP comparison has three more ex-post scored issues per market; that denominator difference is disclosed. The full annual conditional rates for both new arms and both original-UP weighting schemes are in the generated calibration table below.

The six final lane certificates have 474–476 observations in China and 550–551 in the US. Their Pinsker discrepancy allowances range from .1217 to .1221 and from .1140 to .1142, respectively. Thus a measured miss rate above .1 is compatible with the bound. The portfolio uses the clipped positive radius only in its utility, but the miss event and betting wealth use the signed radius. No return or tail-loss guarantee follows from these coverage certificates.

Paired financial uncertainty. The predeclared 20-session circular block bootstrap draws 2,000 paired replicates within each market. For the two-market descriptive pool, draws are independent

between markets and weighted by each market's number of days. All annual and doubled-cost contrasts, including losses against Mean or Fixed, appear in the generated tables. Risk intervals use the same paired blocks and no multiplicity correction. Maximum-drawdown intervals should be read as exploratory path resampling, not a guarantee for future drawdowns.

Table 37: All replay arms in both public markets at base and doubled fees, January 2015–July 2020. Net returns are basis points per day; ES and maximum drawdown are percentages.

Market	Cost	Score	Arm	Net	$ES_{.05}$	MDD
China	1	All	Fixed	2.16	4.41	61.08
China	1	All	IUP	2.61	4.45	58.93
China	1	All	Mean	3.66	4.45	56.85
China	1	All	UP	3.35	4.32	53.21
China	1	All	UUP	2.86	4.38	57.49
China	1	Family	Fixed	5.04	4.18	47.74
China	1	Family	IUP	3.23	4.10	58.28
China	1	Family	Mean	3.92	4.26	55.63
China	1	Family	UP	4.07	4.35	54.92
China	1	Family	UUP	2.90	4.10	58.79
China	2	All	Fixed	0.93	4.42	63.73
China	2	All	IUP	1.52	4.46	61.05
China	2	All	Mean	2.42	4.46	59.88
China	2	All	UP	2.25	4.33	55.76
China	2	All	UUP	1.74	4.39	59.80
China	2	Family	Fixed	4.03	4.19	49.94
China	2	Family	IUP	2.30	4.10	59.66
China	2	Family	Mean	2.90	4.27	57.91
China	2	Family	UP	2.97	4.35	57.34
China	2	Family	UUP	1.96	4.10	60.53
US	1	All	Fixed	2.12	3.56	51.95
US	1	All	IUP	2.83	3.29	49.04
US	1	All	Mean	2.93	3.57	51.97
US	1	All	UP	2.48	3.44	51.87
US	1	All	UUP	2.73	3.27	49.04
US	1	Family	Fixed	2.52	3.27	47.68
US	1	Family	IUP	2.74	3.16	47.59
US	1	Family	Mean	2.73	3.33	49.98
US	1	Family	UP	2.66	3.21	49.36
US	1	Family	UUP	2.66	3.17	47.91
US	2	All	Fixed	0.27	3.58	53.31
US	2	All	IUP	1.00	3.30	49.34
US	2	All	Mean	1.08	3.58	52.25
US	2	All	UP	0.63	3.46	52.15
US	2	All	UUP	0.90	3.29	49.34
US	2	Family	Fixed	0.68	3.29	49.90
US	2	Family	IUP	0.81	3.18	50.16
US	2	Family	Mean	0.88	3.34	50.28
US	2	Family	UP	0.82	3.23	49.95
US	2	Family	UUP	0.74	3.19	50.45

Table 38: Complete exploratory IUP and UUP paired net-return contrasts. Window All is January 2015–July 2020; 2020 is partial. Differences and 95% intervals are basis points per day, without multiplicity correction.

Market	Cost	Window	Score	Arm	vs.	Net difference [95%]
China	1	2015	All	IUP	Fixed	-0.80 [-5.41,+3.36]
China	1	2015	All	IUP	Mean	-0.95 [-6.70,+3.37]
China	1	2015	All	IUP	UP	-2.31 [-7.83,+2.47]

Continued on next page

Table 38 (continued)

Market	Cost	Window	Score	Arm	vs.	Net difference [95%]
China	1	2015	All	UUP	Fixed	+1.16 [-3.43,+6.40]
China	1	2015	All	UUP	Mean	+1.00 [-4.27,+6.19]
China	1	2015	All	UUP	UP	-0.36 [-4.27,+3.30]
China	1	2015	Family	IUP	Fixed	-6.13 [-15.92,+2.12]
China	1	2015	Family	IUP	Mean	-4.38 [-14.99,+2.90]
China	1	2015	Family	IUP	UP	-4.96 [-13.64,+2.49]
China	1	2015	Family	UUP	Fixed	-5.64 [-15.13,+2.29]
China	1	2015	Family	UUP	Mean	-3.89 [-14.38,+3.15]
China	1	2015	Family	UUP	UP	-4.47 [-12.50,+2.49]
China	1	2016	All	IUP	Fixed	-1.08 [-2.45,+0.16]
China	1	2016	All	IUP	Mean	+0.53 [-0.69,+1.96]
China	1	2016	All	IUP	UP	-0.19 [-1.50,+1.01]
China	1	2016	All	UUP	Fixed	-1.05 [-2.45,+0.33]
China	1	2016	All	UUP	Mean	+0.57 [-0.93,+2.31]
China	1	2016	All	UUP	UP	-0.15 [-1.65,+1.33]
China	1	2016	Family	IUP	Fixed	-1.13 [-2.67,+0.17]
China	1	2016	Family	IUP	Mean	-1.35 [-3.17,+0.54]
China	1	2016	Family	IUP	UP	+0.92 [-2.11,+4.45]
China	1	2016	Family	UUP	Fixed	-1.04 [-2.51,+0.14]
China	1	2016	Family	UUP	Mean	-1.26 [-3.14,+0.59]
China	1	2016	Family	UUP	UP	+1.01 [-1.98,+4.40]
China	1	2017	All	IUP	Fixed	+1.81 [-1.61,+5.28]
China	1	2017	All	IUP	Mean	-0.12 [-4.61,+4.49]
China	1	2017	All	IUP	UP	+1.58 [-0.93,+4.54]
China	1	2017	All	UUP	Fixed	+1.55 [-2.35,+5.56]
China	1	2017	All	UUP	Mean	-0.38 [-5.13,+4.31]
China	1	2017	All	UUP	UP	+1.33 [-1.28,+4.21]
China	1	2017	Family	IUP	Fixed	+0.56 [-1.73,+2.60]
China	1	2017	Family	IUP	Mean	+1.46 [-0.53,+3.21]
China	1	2017	Family	IUP	UP	-0.48 [-4.36,+3.22]
China	1	2017	Family	UUP	Fixed	+0.97 [-1.46,+3.20]
China	1	2017	Family	UUP	Mean	+1.87 [-0.27,+3.83]
China	1	2017	Family	UUP	UP	-0.07 [-4.37,+3.99]
China	1	2018	All	IUP	Fixed	+1.23 [-4.74,+8.24]
China	1	2018	All	IUP	Mean	-1.89 [-6.66,+3.25]
China	1	2018	All	IUP	UP	-3.62 [-7.56,+0.38]
China	1	2018	All	UUP	Fixed	+1.19 [-4.88,+7.60]
China	1	2018	All	UUP	Mean	-1.92 [-7.28,+3.45]
China	1	2018	All	UUP	UP	-3.66 [-7.28,+0.03]
China	1	2018	Family	IUP	Fixed	-2.04 [-7.09,+3.01]
China	1	2018	Family	IUP	Mean	+1.48 [-1.16,+4.01]
China	1	2018	Family	IUP	UP	-1.05 [-7.12,+5.35]
China	1	2018	Family	UUP	Fixed	-4.00 [-9.34,+0.78]
China	1	2018	Family	UUP	Mean	-0.48 [-2.70,+1.43]
China	1	2018	Family	UUP	UP	-3.01 [-9.46,+3.31]
China	1	2019	All	IUP	Fixed	+1.06 [-0.98,+3.04]
China	1	2019	All	IUP	Mean	-0.40 [-2.26,+1.26]
China	1	2019	All	IUP	UP	+0.48 [-1.06,+2.07]
China	1	2019	All	UUP	Fixed	+0.66 [-1.32,+2.45]
China	1	2019	All	UUP	Mean	-0.80 [-2.88,+1.00]
China	1	2019	All	UUP	UP	+0.08 [-1.40,+1.65]
China	1	2019	Family	IUP	Fixed	-0.64 [-2.46,+1.44]
China	1	2019	Family	IUP	Mean	-0.72 [-3.49,+2.03]
China	1	2019	Family	IUP	UP	-0.33 [-1.97,+1.37]
China	1	2019	Family	UUP	Fixed	-0.75 [-2.58,+1.30]
China	1	2019	Family	UUP	Mean	-0.83 [-3.54,+1.86]
China	1	2019	Family	UUP	UP	-0.44 [-1.93,+1.15]
China	1	2020	All	IUP	Fixed	+0.15 [-1.76,+2.06]
China	1	2020	All	IUP	Mean	-5.04 [-7.72,-1.59]
China	1	2020	All	IUP	UP	-0.24 [-1.77,+1.37]
China	1	2020	All	UUP	Fixed	+0.46 [-1.97,+2.95]

Continued on next page

Table 38 (continued)

Market	Cost	Window	Score	Arm	vs.	Net difference [95%]
China	1	2020	All	UUP	Mean	-4.74 [-7.83,-0.63]
China	1	2020	All	UUP	UP	+0.06 [-2.08,+2.20]
China	1	2020	Family	IUP	Fixed	-1.64 [-4.10,+0.82]
China	1	2020	Family	IUP	Mean	-1.07 [-4.91,+3.80]
China	1	2020	Family	IUP	UP	+1.89 [-0.12,+3.90]
China	1	2020	Family	UUP	Fixed	-2.79 [-5.26,-0.24]
China	1	2020	Family	UUP	Mean	-2.22 [-6.47,+2.69]
China	1	2020	Family	UUP	UP	+0.73 [-1.25,+2.80]
China	1	All	All	IUP	Fixed	+0.45 [-1.14,+2.03]
China	1	All	All	IUP	Mean	-1.05 [-2.69,+0.53]
China	1	All	All	IUP	UP	-0.74 [-2.13,+0.54]
China	1	All	All	UUP	Fixed	+0.70 [-0.96,+2.47]
China	1	All	All	UUP	Mean	-0.80 [-2.62,+0.92]
China	1	All	All	UUP	UP	-0.49 [-1.76,+0.66]
China	1	All	Family	IUP	Fixed	-1.81 [-3.96,+0.01]
China	1	All	Family	IUP	Mean	-0.69 [-2.75,+0.96]
China	1	All	Family	IUP	UP	-0.84 [-3.01,+1.31]
China	1	All	Family	UUP	Fixed	-2.14 [-4.28,-0.35]
China	1	All	Family	UUP	Mean	-1.02 [-3.01,+0.59]
China	1	All	Family	UUP	UP	-1.17 [-3.34,+0.93]
China	2	2015	All	IUP	Fixed	-0.84 [-5.51,+3.32]
China	2	2015	All	IUP	Mean	-0.98 [-6.77,+3.39]
China	2	2015	All	IUP	UP	-2.32 [-7.87,+2.47]
China	2	2015	All	UUP	Fixed	+1.03 [-3.54,+6.13]
China	2	2015	All	UUP	Mean	+0.89 [-4.37,+6.03]
China	2	2015	All	UUP	UP	-0.45 [-4.38,+3.18]
China	2	2015	Family	IUP	Fixed	-5.87 [-15.70,+2.30]
China	2	2015	Family	IUP	Mean	-3.81 [-14.46,+3.44]
China	2	2015	Family	IUP	UP	-4.92 [-13.57,+2.53]
China	2	2015	Family	UUP	Fixed	-5.39 [-14.86,+2.43]
China	2	2015	Family	UUP	Mean	-3.33 [-13.79,+3.69]
China	2	2015	Family	UUP	UP	-4.44 [-12.46,+2.56]
China	2	2016	All	IUP	Fixed	-1.07 [-2.44,+0.17]
China	2	2016	All	IUP	Mean	+0.52 [-0.69,+1.93]
China	2	2016	All	IUP	UP	-0.10 [-1.42,+1.16]
China	2	2016	All	UUP	Fixed	-1.04 [-2.44,+0.34]
China	2	2016	All	UUP	Mean	+0.55 [-0.93,+2.28]
China	2	2016	All	UUP	UP	-0.07 [-1.60,+1.42]
China	2	2016	Family	IUP	Fixed	-1.36 [-2.90,-0.05]
China	2	2016	Family	IUP	Mean	-1.41 [-3.22,+0.49]
China	2	2016	Family	IUP	UP	+0.71 [-2.34,+4.15]
China	2	2016	Family	UUP	Fixed	-1.25 [-2.75,-0.05]
China	2	2016	Family	UUP	Mean	-1.30 [-3.16,+0.56]
China	2	2016	Family	UUP	UP	+0.81 [-2.19,+4.09]
China	2	2017	All	IUP	Fixed	+1.88 [-1.55,+5.37]
China	2	2017	All	IUP	Mean	-0.01 [-4.49,+4.61]
China	2	2017	All	IUP	UP	+1.55 [-0.92,+4.46]
China	2	2017	All	UUP	Fixed	+1.64 [-2.27,+5.62]
China	2	2017	All	UUP	Mean	-0.25 [-5.00,+4.43]
China	2	2017	All	UUP	UP	+1.32 [-1.28,+4.17]
China	2	2017	Family	IUP	Fixed	+0.49 [-1.78,+2.52]
China	2	2017	Family	IUP	Mean	+1.40 [-0.59,+3.12]
China	2	2017	Family	IUP	UP	-0.25 [-4.23,+3.44]
China	2	2017	Family	UUP	Fixed	+0.90 [-1.49,+3.13]
China	2	2017	Family	UUP	Mean	+1.81 [-0.31,+3.78]
China	2	2017	Family	UUP	UP	+0.16 [-4.17,+4.24]
China	2	2018	All	IUP	Fixed	+1.97 [-4.00,+9.01]
China	2	2018	All	IUP	Mean	-1.11 [-5.88,+4.15]
China	2	2018	All	IUP	UP	-3.59 [-7.57,+0.47]
China	2	2018	All	UUP	Fixed	+1.83 [-4.29,+8.25]
China	2	2018	All	UUP	Mean	-1.25 [-6.52,+4.25]

Continued on next page

Table 38 (continued)

Market	Cost	Window	Score	Arm	vs.	Net difference [95%]
China	2	2018	All	UUP	UP	-3.73 [-7.31,-0.06]
China	2	2018	Family	IUP	Fixed	-1.54 [-6.62,+3.55]
China	2	2018	Family	IUP	Mean	+1.59 [-1.06,+4.15]
China	2	2018	Family	IUP	UP	-0.16 [-6.31,+6.31]
China	2	2018	Family	UUP	Fixed	-3.57 [-8.90,+1.21]
China	2	2018	Family	UUP	Mean	-0.45 [-2.68,+1.49]
China	2	2018	Family	UUP	UP	-2.20 [-8.56,+4.16]
China	2	2019	All	IUP	Fixed	+1.06 [-0.99,+3.06]
China	2	2019	All	IUP	Mean	-0.41 [-2.27,+1.24]
China	2	2019	All	IUP	UP	+0.47 [-1.08,+2.05]
China	2	2019	All	UUP	Fixed	+0.66 [-1.32,+2.47]
China	2	2019	All	UUP	Mean	-0.81 [-2.91,+0.99]
China	2	2019	All	UUP	UP	+0.07 [-1.41,+1.64]
China	2	2019	Family	IUP	Fixed	-0.66 [-2.47,+1.42]
China	2	2019	Family	IUP	Mean	-0.75 [-3.52,+1.99]
China	2	2019	Family	IUP	UP	-0.34 [-1.98,+1.34]
China	2	2019	Family	UUP	Fixed	-0.77 [-2.59,+1.27]
China	2	2019	Family	UUP	Mean	-0.86 [-3.53,+1.81]
China	2	2019	Family	UUP	UP	-0.45 [-1.98,+1.12]
China	2	2020	All	IUP	Fixed	+0.17 [-1.74,+2.09]
China	2	2020	All	IUP	Mean	-5.03 [-7.71,-1.58]
China	2	2020	All	IUP	UP	-0.22 [-1.75,+1.38]
China	2	2020	All	UUP	Fixed	+0.46 [-1.98,+2.96]
China	2	2020	All	UUP	Mean	-4.75 [-7.83,-0.64]
China	2	2020	All	UUP	UP	+0.06 [-2.09,+2.20]
China	2	2020	Family	IUP	Fixed	-1.62 [-4.09,+0.87]
China	2	2020	Family	IUP	Mean	-1.05 [-4.90,+3.81]
China	2	2020	Family	IUP	UP	+1.91 [-0.10,+3.91]
China	2	2020	Family	UUP	Fixed	-2.77 [-5.25,-0.22]
China	2	2020	Family	UUP	Mean	-2.21 [-6.45,+2.71]
China	2	2020	Family	UUP	UP	+0.75 [-1.23,+2.80]
China	2	All	All	IUP	Fixed	+0.59 [-1.02,+2.22]
China	2	All	All	IUP	Mean	-0.89 [-2.55,+0.70]
China	2	All	All	IUP	UP	-0.73 [-2.13,+0.57]
China	2	All	All	UUP	Fixed	+0.81 [-0.82,+2.57]
China	2	All	All	UUP	Mean	-0.68 [-2.51,+1.05]
China	2	All	All	UUP	UP	-0.51 [-1.77,+0.62]
China	2	All	Family	IUP	Fixed	-1.72 [-3.84,+0.11]
China	2	All	Family	IUP	Mean	-0.60 [-2.62,+1.04]
China	2	All	Family	IUP	UP	-0.66 [-2.86,+1.47]
China	2	All	Family	UUP	Fixed	-2.07 [-4.19,-0.28]
China	2	All	Family	UUP	Mean	-0.94 [-2.92,+0.67]
China	2	All	Family	UUP	UP	-1.00 [-3.14,+1.10]
US	1	2015	All	IUP	Fixed	+1.38 [-1.32,+4.18]
US	1	2015	All	IUP	Mean	+0.62 [-1.31,+2.84]
US	1	2015	All	IUP	UP	+1.04 [-0.80,+3.15]
US	1	2015	All	UUP	Fixed	+1.35 [-1.33,+4.21]
US	1	2015	All	UUP	Mean	+0.59 [-1.33,+2.84]
US	1	2015	All	UUP	UP	+1.02 [-0.80,+3.09]
US	1	2015	Family	IUP	Fixed	+0.70 [-1.96,+3.74]
US	1	2015	Family	IUP	Mean	+0.59 [-0.87,+2.31]
US	1	2015	Family	IUP	UP	+0.12 [-0.94,+1.29]
US	1	2015	Family	UUP	Fixed	+0.97 [-1.69,+3.96]
US	1	2015	Family	UUP	Mean	+0.86 [-0.64,+2.55]
US	1	2015	Family	UUP	UP	+0.39 [-0.71,+1.58]
US	1	2016	All	IUP	Fixed	-1.58 [-4.96,+1.73]
US	1	2016	All	IUP	Mean	-1.83 [-4.83,+1.23]
US	1	2016	All	IUP	UP	-1.63 [-3.28,+0.14]
US	1	2016	All	UUP	Fixed	-1.67 [-5.07,+1.78]
US	1	2016	All	UUP	Mean	-1.93 [-4.87,+1.18]
US	1	2016	All	UUP	UP	-1.72 [-3.39,+0.18]

Continued on next page

Table 38 (continued)

Market	Cost	Window	Score	Arm	vs.	Net difference [95%]
US	1	2016	Family	IUP	Fixed	-0.13 [-3.06,+3.04]
US	1	2016	Family	IUP	Mean	-1.70 [-4.23,+0.58]
US	1	2016	Family	IUP	UP	-0.06 [-1.19,+1.10]
US	1	2016	Family	UUP	Fixed	-0.24 [-3.24,+3.04]
US	1	2016	Family	UUP	Mean	-1.81 [-4.47,+0.61]
US	1	2016	Family	UUP	UP	-0.17 [-1.32,+1.04]
US	1	2017	All	IUP	Fixed	+0.27 [-1.76,+2.43]
US	1	2017	All	IUP	Mean	-0.48 [-2.16,+1.23]
US	1	2017	All	IUP	UP	+0.18 [-1.24,+1.67]
US	1	2017	All	UUP	Fixed	+0.17 [-1.76,+2.23]
US	1	2017	All	UUP	Mean	-0.58 [-2.29,+1.10]
US	1	2017	All	UUP	UP	+0.08 [-1.18,+1.43]
US	1	2017	Family	IUP	Fixed	+2.14 [+0.40,+3.88]
US	1	2017	Family	IUP	Mean	+1.00 [-0.45,+2.33]
US	1	2017	Family	IUP	UP	-0.42 [-1.47,+0.53]
US	1	2017	Family	UUP	Fixed	+2.01 [+0.23,+3.80]
US	1	2017	Family	UUP	Mean	+0.87 [-0.66,+2.32]
US	1	2017	Family	UUP	UP	-0.55 [-1.67,+0.51]
US	1	2018	All	IUP	Fixed	+1.69 [-0.30,+3.51]
US	1	2018	All	IUP	Mean	+0.85 [-0.25,+1.95]
US	1	2018	All	IUP	UP	+0.20 [-0.87,+1.32]
US	1	2018	All	UUP	Fixed	+1.39 [-0.57,+3.26]
US	1	2018	All	UUP	Mean	+0.55 [-0.54,+1.59]
US	1	2018	All	UUP	UP	-0.10 [-1.08,+0.91]
US	1	2018	Family	IUP	Fixed	+0.47 [-1.27,+2.10]
US	1	2018	Family	IUP	Mean	-0.66 [-2.29,+0.94]
US	1	2018	Family	IUP	UP	-0.56 [-1.58,+0.60]
US	1	2018	Family	UUP	Fixed	+0.69 [-1.13,+2.35]
US	1	2018	Family	UUP	Mean	-0.44 [-2.00,+1.22]
US	1	2018	Family	UUP	UP	-0.34 [-1.37,+0.76]
US	1	2019	All	IUP	Fixed	+0.55 [-0.92,+1.88]
US	1	2019	All	IUP	Mean	+0.61 [-1.41,+2.72]
US	1	2019	All	IUP	UP	+0.49 [-1.38,+1.95]
US	1	2019	All	UUP	Fixed	+0.61 [-0.93,+2.01]
US	1	2019	All	UUP	Mean	+0.67 [-1.46,+2.94]
US	1	2019	All	UUP	UP	+0.55 [-1.31,+1.99]
US	1	2019	Family	IUP	Fixed	-1.84 [-3.71,-0.12]
US	1	2019	Family	IUP	Mean	-1.03 [-2.62,+0.68]
US	1	2019	Family	IUP	UP	-0.55 [-2.01,+0.84]
US	1	2019	Family	UUP	Fixed	-1.99 [-4.01,-0.05]
US	1	2019	Family	UUP	Mean	-1.19 [-2.79,+0.57]
US	1	2019	Family	UUP	UP	-0.71 [-2.33,+0.84]
US	1	2020	All	IUP	Fixed	+2.89 [-2.36,+10.32]
US	1	2020	All	IUP	Mean	-0.49 [-6.58,+7.11]
US	1	2020	All	IUP	UP	+2.92 [-2.12,+9.69]
US	1	2020	All	UUP	Fixed	+2.73 [-2.71,+10.32]
US	1	2020	All	UUP	Mean	-0.65 [-6.78,+7.07]
US	1	2020	All	UUP	UP	+2.76 [-2.44,+9.66]
US	1	2020	Family	IUP	Fixed	-0.17 [-4.15,+3.53]
US	1	2020	Family	IUP	Mean	+3.23 [-1.97,+8.74]
US	1	2020	Family	IUP	UP	+3.30 [-0.11,+7.45]
US	1	2020	Family	UUP	Fixed	-1.06 [-5.49,+2.91]
US	1	2020	Family	UUP	Mean	+2.34 [-3.10,+7.77]
US	1	2020	Family	UUP	UP	+2.41 [-0.66,+5.91]
US	1	All	All	IUP	Fixed	+0.71 [-0.44,+1.99]
US	1	All	All	IUP	Mean	-0.09 [-1.15,+1.08]
US	1	All	All	IUP	UP	+0.36 [-0.50,+1.38]
US	1	All	All	UUP	Fixed	+0.61 [-0.54,+1.91]
US	1	All	All	UUP	Mean	-0.19 [-1.27,+1.01]
US	1	All	All	UUP	UP	+0.26 [-0.59,+1.28]
US	1	All	Family	IUP	Fixed	+0.22 [-0.79,+1.23]

Continued on next page

Table 38 (continued)

Market	Cost	Window	Score	Arm	vs.	Net difference [95%]
US	1	All	Family	IUP	Mean	+0.01 [-0.92,+0.95]
US	1	All	Family	IUP	UP	+0.08 [-0.53,+0.76]
US	1	All	Family	UUP	Fixed	+0.14 [-0.92,+1.25]
US	1	All	Family	UUP	Mean	-0.06 [-1.03,+0.86]
US	1	All	Family	UUP	UP	+0.00 [-0.59,+0.66]
US	2	2015	All	IUP	Fixed	+1.39 [-1.31,+4.19]
US	2	2015	All	IUP	Mean	+0.60 [-1.35,+2.83]
US	2	2015	All	IUP	UP	+1.04 [-0.80,+3.16]
US	2	2015	All	UUP	Fixed	+1.37 [-1.32,+4.22]
US	2	2015	All	UUP	Mean	+0.58 [-1.35,+2.82]
US	2	2015	All	UUP	UP	+1.02 [-0.82,+3.09]
US	2	2015	Family	IUP	Fixed	+0.71 [-1.95,+3.75]
US	2	2015	Family	IUP	Mean	+0.58 [-0.88,+2.31]
US	2	2015	Family	IUP	UP	+0.12 [-0.93,+1.29]
US	2	2015	Family	UUP	Fixed	+0.98 [-1.68,+3.98]
US	2	2015	Family	UUP	Mean	+0.85 [-0.65,+2.54]
US	2	2015	Family	UUP	UP	+0.39 [-0.71,+1.58]
US	2	2016	All	IUP	Fixed	-1.58 [-4.96,+1.73]
US	2	2016	All	IUP	Mean	-1.83 [-4.82,+1.24]
US	2	2016	All	IUP	UP	-1.63 [-3.27,+0.15]
US	2	2016	All	UUP	Fixed	-1.67 [-5.07,+1.78]
US	2	2016	All	UUP	Mean	-1.92 [-4.87,+1.18]
US	2	2016	All	UUP	UP	-1.72 [-3.38,+0.18]
US	2	2016	Family	IUP	Fixed	-0.13 [-3.05,+3.04]
US	2	2016	Family	IUP	Mean	-1.69 [-4.21,+0.59]
US	2	2016	Family	IUP	UP	-0.06 [-1.20,+1.09]
US	2	2016	Family	UUP	Fixed	-0.24 [-3.23,+3.04]
US	2	2016	Family	UUP	Mean	-1.81 [-4.46,+0.62]
US	2	2016	Family	UUP	UP	-0.18 [-1.32,+1.03]
US	2	2017	All	IUP	Fixed	+0.28 [-1.74,+2.45]
US	2	2017	All	IUP	Mean	-0.46 [-2.15,+1.25]
US	2	2017	All	IUP	UP	+0.19 [-1.24,+1.68]
US	2	2017	All	UUP	Fixed	+0.18 [-1.74,+2.24]
US	2	2017	All	UUP	Mean	-0.57 [-2.27,+1.11]
US	2	2017	All	UUP	UP	+0.08 [-1.17,+1.43]
US	2	2017	Family	IUP	Fixed	+2.15 [+0.40,+3.88]
US	2	2017	Family	IUP	Mean	+1.00 [-0.44,+2.34]
US	2	2017	Family	IUP	UP	-0.42 [-1.47,+0.53]
US	2	2017	Family	UUP	Fixed	+2.02 [+0.23,+3.81]
US	2	2017	Family	UUP	Mean	+0.88 [-0.65,+2.32]
US	2	2017	Family	UUP	UP	-0.55 [-1.67,+0.51]
US	2	2018	All	IUP	Fixed	+1.69 [-0.29,+3.51]
US	2	2018	All	IUP	Mean	+0.85 [-0.24,+1.95]
US	2	2018	All	IUP	UP	+0.20 [-0.87,+1.32]
US	2	2018	All	UUP	Fixed	+1.39 [-0.57,+3.27]
US	2	2018	All	UUP	Mean	+0.55 [-0.55,+1.59]
US	2	2018	All	UUP	UP	-0.10 [-1.07,+0.92]
US	2	2018	Family	IUP	Fixed	+0.47 [-1.27,+2.11]
US	2	2018	Family	IUP	Mean	-0.66 [-2.28,+0.94]
US	2	2018	Family	IUP	UP	-0.56 [-1.57,+0.60]
US	2	2018	Family	UUP	Fixed	+0.69 [-1.12,+2.35]
US	2	2018	Family	UUP	Mean	-0.44 [-2.00,+1.22]
US	2	2018	Family	UUP	UP	-0.34 [-1.37,+0.76]
US	2	2019	All	IUP	Fixed	+0.55 [-0.92,+1.88]
US	2	2019	All	IUP	Mean	+0.61 [-1.41,+2.72]
US	2	2019	All	IUP	UP	+0.49 [-1.38,+1.95]
US	2	2019	All	UUP	Fixed	+0.61 [-0.93,+2.01]
US	2	2019	All	UUP	Mean	+0.67 [-1.46,+2.94]
US	2	2019	All	UUP	UP	+0.55 [-1.30,+1.98]
US	2	2019	Family	IUP	Fixed	-2.17 [-4.03,-0.48]
US	2	2019	Family	IUP	Mean	-1.37 [-2.97,+0.36]

Continued on next page

Table 38 (continued)

Market	Cost	Window	Score	Arm	vs.	Net difference [95%]
US	2	2019	Family	IUP	UP	-0.89 [-2.36,+0.51]
US	2	2019	Family	UUP	Fixed	-2.33 [-4.31,-0.41]
US	2	2019	Family	UUP	Mean	-1.52 [-3.13,+0.24]
US	2	2019	Family	UUP	UP	-1.05 [-2.66,+0.53]
US	2	2020	All	IUP	Fixed	+3.02 [-2.18,+10.35]
US	2	2020	All	IUP	Mean	-0.36 [-6.33,+7.18]
US	2	2020	All	IUP	UP	+3.04 [-2.00,+9.74]
US	2	2020	All	UUP	Fixed	+2.87 [-2.55,+10.35]
US	2	2020	All	UUP	Mean	-0.51 [-6.58,+7.15]
US	2	2020	All	UUP	UP	+2.89 [-2.25,+9.75]
US	2	2020	Family	IUP	Fixed	-0.41 [-4.43,+3.31]
US	2	2020	Family	IUP	Mean	+2.99 [-2.16,+8.45]
US	2	2020	Family	IUP	UP	+3.06 [-0.33,+7.14]
US	2	2020	Family	UUP	Fixed	-1.30 [-5.80,+2.70]
US	2	2020	Family	UUP	Mean	+2.10 [-3.34,+7.53]
US	2	2020	Family	UUP	UP	+2.17 [-0.89,+5.61]
US	2	All	All	IUP	Fixed	+0.73 [-0.42,+2.00]
US	2	All	All	IUP	Mean	-0.08 [-1.14,+1.10]
US	2	All	All	IUP	UP	+0.37 [-0.48,+1.39]
US	2	All	All	UUP	Fixed	+0.63 [-0.51,+1.92]
US	2	All	All	UUP	Mean	-0.18 [-1.25,+1.02]
US	2	All	All	UUP	UP	+0.27 [-0.57,+1.29]
US	2	All	Family	IUP	Fixed	+0.14 [-0.87,+1.16]
US	2	All	Family	IUP	Mean	-0.07 [-1.00,+0.86]
US	2	All	Family	IUP	UP	-0.00 [-0.62,+0.68]
US	2	All	Family	UUP	Fixed	+0.06 [-1.01,+1.18]
US	2	All	Family	UUP	Mean	-0.15 [-1.12,+0.79]
US	2	All	Family	UUP	UP	-0.08 [-0.67,+0.57]

Table 39: Observed-label conditional fractional misses and original UP under issue-time sensitivity/uniform weights. Days counts observed feedback issues; Mass is mean observed weight, H_t. UP (influence) and UP (uniform) evaluate factorwise UP using the indicated weights. The original-UP audit includes three additional terminal issues per market; a same-feedback audit is reported in text.

Market	Window	Method	Days	Miss (%)	Mass (%)
China	2015	IUP	217	14.83	13.98
China	2015	UP (influence)	217	23.39	13.98
China	2015	UP (uniform)	217	23.06	14.36
China	2015	UUP	217	14.47	14.36
China	2016	IUP	223	13.89	8.00
China	2016	UP (influence)	223	14.86	8.00
China	2016	UP (uniform)	223	14.45	8.87
China	2016	UUP	223	13.79	8.87
China	2017	IUP	165	14.45	2.48
China	2017	UP (influence)	165	13.82	2.48
China	2017	UP (uniform)	165	13.60	2.75
China	2017	UUP	165	14.48	2.75
China	2018	IUP	198	13.20	8.98
China	2018	UP (influence)	198	21.60	8.98
China	2018	UP (uniform)	198	21.58	9.95
China	2018	UUP	198	13.28	9.95
China	2019	IUP	241	13.40	68.53
China	2019	UP (influence)	241	15.72	68.53
China	2019	UP (uniform)	241	15.54	68.34
China	2019	UUP	241	13.45	68.34
China	2020	IUP	137	14.09	81.81
China	2020	UP (influence)	140	16.06	81.85

Continued on next page

Table 39 (continued)

Market	Window	Method	Days	Miss (%)	Mass (%)
China	2020	UP (uniform)	140	15.93	78.01
China	2020	UUP	137	14.28	77.98
China	All	IUP	1181	13.95	29.41
China	All	UP (influence)	1184	17.72	29.54
China	All	UP (uniform)	1184	17.50	29.48
China	All	UUP	1181	13.91	29.36
US	2015	IUP	250	14.44	73.33
US	2015	UP (influence)	250	15.23	73.33
US	2015	UP (uniform)	250	15.11	77.46
US	2015	UUP	250	14.43	77.46
US	2016	IUP	252	12.93	84.88
US	2016	UP (influence)	252	12.35	84.88
US	2016	UP (uniform)	252	12.35	85.82
US	2016	UUP	252	12.86	85.82
US	2017	IUP	251	12.69	91.93
US	2017	UP (influence)	251	12.80	91.93
US	2017	UP (uniform)	251	12.88	91.29
US	2017	UUP	251	12.77	91.29
US	2018	IUP	251	13.49	68.06
US	2018	UP (influence)	251	15.72	68.06
US	2018	UP (uniform)	251	15.80	68.52
US	2018	UUP	251	13.53	68.52
US	2019	IUP	252	11.36	26.09
US	2019	UP (influence)	252	9.74	26.09
US	2019	UP (uniform)	252	9.87	28.09
US	2019	UUP	252	11.38	28.09
US	2020	IUP	144	12.60	23.81
US	2020	UP (influence)	147	17.68	24.05
US	2020	UP (uniform)	147	17.87	23.72
US	2020	UUP	144	12.57	23.73
US	All	IUP	1400	12.94	64.20
US	All	UP (influence)	1403	13.63	64.14
US	All	UP (uniform)	1403	13.69	65.34
US	All	UUP	1400	12.95	65.43

Table 40: All exploratory paired risk comparisons for the full 2015–2020 period. Negative percentage-point differences mean lower loss; intervals are unadjusted 20-session block-bootstrap diagnostics.

Market	Cost	Score	Arm	vs.	Metric	Diff. [95%] pp
China	1	All	IUP	Fixed	$ES_{.05}$	+0.043 [-0.144,+0.242]
China	1	All	IUP	Fixed	MDD	-2.152 [-8.001,+5.775]
China	1	All	IUP	Mean	$ES_{.05}$	-0.000 [-0.215,+0.217]
China	1	All	IUP	Mean	MDD	+2.081 [-3.662,+9.789]
China	1	All	IUP	UP	$ES_{.05}$	+0.126 [-0.043,+0.305]
China	1	All	IUP	UP	MDD	+5.721 [-2.202,+8.913]
China	1	All	UUP	Fixed	$ES_{.05}$	-0.024 [-0.179,+0.151]
China	1	All	UUP	Fixed	MDD	-3.596 [-9.323,+4.324]
China	1	All	UUP	Mean	$ES_{.05}$	-0.067 [-0.247,+0.119]
China	1	All	UUP	Mean	MDD	+0.637 [-5.429,+8.557]
China	1	All	UUP	UP	$ES_{.05}$	+0.059 [-0.069,+0.186]
China	1	All	UUP	UP	MDD	+4.276 [-2.763,+7.184]
China	1	Family	IUP	Fixed	$ES_{.05}$	-0.085 [-0.257,+0.088]
China	1	Family	IUP	Fixed	MDD	+10.539 [-3.018,+12.701]
China	1	Family	IUP	Mean	$ES_{.05}$	-0.159 [-0.403,+0.041]
China	1	Family	IUP	Mean	MDD	+2.646 [-5.153,+9.609]
China	1	Family	IUP	UP	$ES_{.05}$	-0.246 [-0.499,-0.026]
China	1	Family	IUP	UP	MDD	+3.357 [-7.123,+6.928]
China	1	Family	UUP	Fixed	$ES_{.05}$	-0.086 [-0.252,+0.088]
China	1	Family	UUP	Fixed	MDD	+11.053 [-2.137,+13.866]
China	1	Family	UUP	Mean	$ES_{.05}$	-0.160 [-0.394,+0.047]
China	1	Family	UUP	Mean	MDD	+3.160 [-4.046,+10.361]
China	1	Family	UUP	UP	$ES_{.05}$	-0.247 [-0.500,-0.026]
China	1	Family	UUP	UP	MDD	+3.871 [-5.982,+8.202]
China	2	All	IUP	Fixed	$ES_{.05}$	+0.043 [-0.146,+0.242]
China	2	All	IUP	Fixed	MDD	-2.681 [-8.702,+5.177]
China	2	All	IUP	Mean	$ES_{.05}$	-0.001 [-0.217,+0.217]
China	2	All	IUP	Mean	MDD	+1.173 [-4.293,+9.338]
China	2	All	IUP	UP	$ES_{.05}$	+0.126 [-0.043,+0.306]
China	2	All	IUP	UP	MDD	+5.288 [-2.305,+9.130]
China	2	All	UUP	Fixed	$ES_{.05}$	-0.024 [-0.182,+0.151]
China	2	All	UUP	Fixed	MDD	-3.935 [-9.768,+4.138]
China	2	All	UUP	Mean	$ES_{.05}$	-0.067 [-0.249,+0.120]
China	2	All	UUP	Mean	MDD	-0.081 [-5.582,+8.214]
China	2	All	UUP	UP	$ES_{.05}$	+0.059 [-0.069,+0.188]
China	2	All	UUP	UP	MDD	+4.035 [-2.767,+7.461]
China	2	Family	IUP	Fixed	$ES_{.05}$	-0.087 [-0.259,+0.086]
China	2	Family	IUP	Fixed	MDD	+9.724 [-3.289,+12.968]
China	2	Family	IUP	Mean	$ES_{.05}$	-0.163 [-0.408,+0.038]
China	2	Family	IUP	Mean	MDD	+1.754 [-5.542,+9.440]
China	2	Family	IUP	UP	$ES_{.05}$	-0.248 [-0.501,-0.025]
China	2	Family	IUP	UP	MDD	+2.326 [-7.698,+6.752]
China	2	Family	UUP	Fixed	$ES_{.05}$	-0.087 [-0.254,+0.084]
China	2	Family	UUP	Fixed	MDD	+10.590 [-2.100,+13.921]
China	2	Family	UUP	Mean	$ES_{.05}$	-0.164 [-0.399,+0.044]
China	2	Family	UUP	Mean	MDD	+2.620 [-4.408,+9.942]
China	2	Family	UUP	UP	$ES_{.05}$	-0.248 [-0.502,-0.026]
China	2	Family	UUP	UP	MDD	+3.192 [-6.486,+7.827]
US	1	All	IUP	Fixed	$ES_{.05}$	-0.279 [-0.443,-0.125]
US	1	All	IUP	Fixed	MDD	-2.908 [-9.213,-0.561]
US	1	All	IUP	Mean	$ES_{.05}$	-0.280 [-0.494,-0.098]
US	1	All	IUP	Mean	MDD	-2.928 [-6.874,+0.395]
US	1	All	IUP	UP	$ES_{.05}$	-0.157 [-0.315,-0.031]
US	1	All	IUP	UP	MDD	-2.836 [-6.334,+0.754]
US	1	All	UUP	Fixed	$ES_{.05}$	-0.291 [-0.463,-0.132]
US	1	All	UUP	Fixed	MDD	-2.910 [-8.839,-0.356]
US	1	All	UUP	Mean	$ES_{.05}$	-0.293 [-0.513,-0.106]
US	1	All	UUP	Mean	MDD	-2.930 [-6.645,+0.686]

Continued on next page

Table 40 (continued)

Market	Cost	Score	Arm	vs.	Metric	Diff. [95%] pp
US	1	All	UUP	UP	$ES_{.05}$	-0.169 [-0.328,-0.042]
US	1	All	UUP	UP	MDD	-2.838 [-5.892,+1.122]
US	1	Family	IUP	Fixed	$ES_{.05}$	-0.114 [-0.227,+0.008]
US	1	Family	IUP	Fixed	MDD	-0.082 [-6.213,+1.935]
US	1	Family	IUP	Mean	$ES_{.05}$	-0.171 [-0.310,-0.058]
US	1	Family	IUP	Mean	MDD	-2.389 [-5.323,+1.378]
US	1	Family	IUP	UP	$ES_{.05}$	-0.056 [-0.114,+0.004]
US	1	Family	IUP	UP	MDD	-1.763 [-3.467,+2.046]
US	1	Family	UUP	Fixed	$ES_{.05}$	-0.101 [-0.219,+0.031]
US	1	Family	UUP	Fixed	MDD	+0.236 [-6.168,+2.274]
US	1	Family	UUP	Mean	$ES_{.05}$	-0.158 [-0.297,-0.051]
US	1	Family	UUP	Mean	MDD	-2.071 [-5.177,+1.488]
US	1	Family	UUP	UP	$ES_{.05}$	-0.043 [-0.103,+0.011]
US	1	Family	UUP	UP	MDD	-1.445 [-3.115,+2.076]
US	2	All	IUP	Fixed	$ES_{.05}$	-0.279 [-0.442,-0.125]
US	2	All	IUP	Fixed	MDD	-3.967 [-9.660,-0.562]
US	2	All	IUP	Mean	$ES_{.05}$	-0.280 [-0.493,-0.098]
US	2	All	IUP	Mean	MDD	-2.904 [-6.834,+0.705]
US	2	All	IUP	UP	$ES_{.05}$	-0.157 [-0.315,-0.031]
US	2	All	IUP	UP	MDD	-2.813 [-6.445,+0.941]
US	2	All	UUP	Fixed	$ES_{.05}$	-0.291 [-0.462,-0.133]
US	2	All	UUP	Fixed	MDD	-3.969 [-9.215,-0.306]
US	2	All	UUP	Mean	$ES_{.05}$	-0.292 [-0.512,-0.106]
US	2	All	UUP	Mean	MDD	-2.906 [-6.556,+1.004]
US	2	All	UUP	UP	$ES_{.05}$	-0.169 [-0.327,-0.043]
US	2	All	UUP	UP	MDD	-2.815 [-6.151,+1.213]
US	2	Family	IUP	Fixed	$ES_{.05}$	-0.110 [-0.224,+0.013]
US	2	Family	IUP	Fixed	MDD	+0.253 [-7.218,+2.186]
US	2	Family	IUP	Mean	$ES_{.05}$	-0.166 [-0.303,-0.056]
US	2	Family	IUP	Mean	MDD	-0.123 [-5.081,+2.067]
US	2	Family	IUP	UP	$ES_{.05}$	-0.052 [-0.108,+0.006]
US	2	Family	IUP	UP	MDD	+0.204 [-3.161,+2.649]
US	2	Family	UUP	Fixed	$ES_{.05}$	-0.098 [-0.218,+0.037]
US	2	Family	UUP	Fixed	MDD	+0.548 [-7.249,+2.629]
US	2	Family	UUP	Mean	$ES_{.05}$	-0.154 [-0.290,-0.048]
US	2	Family	UUP	Mean	MDD	+0.171 [-4.884,+2.362]
US	2	Family	UUP	UP	$ES_{.05}$	-0.039 [-0.098,+0.014]
US	2	Family	UUP	UP	MDD	+0.498 [-2.796,+2.759]