A GENERAL FRAMEWORK FOR BUDGETED THRESHOLD INCENTIVES ON REQUEST

Anonymous authors
Paper under double-blind review

ABSTRACT

On-demand delivery platforms pay riders through incentive activities whose tiers are set from recent completions of riders with a similar history, so that a little extra effort earns a clearly stated reward. Operators request such plans for changing periods, rider populations, payment rules and budgets, days ahead and within minutes, often for holidays or bad weather where randomized trials are scarce and take months to collect. We present a request-driven framework that composes four stages—conditional prediction, population reduction, trajectory integration and budget allocation—through seven replaceable modules that exchange conditional trajectory laws, whose award probabilities and award-marked moments give payment and uplift for any activity rule. To shorten a long randomized campaign, a response-correction step reweights trajectories from abundant no-offer history to match the moments of a short pilot. We prove that, on a fixed plan menu and given the stage errors, the end-to-end value loss is bounded by the sum of four stage terms—synthetic data, moment matching, integration and decision—with constants that cannot be improved from the final tables, and that for every stage there are instances on which omitting it leaves an error floor the others cannot remove. On 3,000 riders over 45 weekly origins, re-drawing trajectories and solving exactly for every request takes 117.8 s per week-long request against 3.07 s for the framework; including the one-off sampling pass, all 127 windows of a week are answered $11.04\times$ faster with identical scenarios and at most 0.92% value lost by the allocation on the audited week-long tables, a gain that comes entirely from reuse, while a point forecast, independent days or an equal budget split each lose accuracy. Holding cities out of training changes plan-cost error by $+0.003$ $[-0.018, +0.020]$, within a pre-set margin. On 24 new controlled response laws, with the plan, the optimum and validity all judged within 20% of the budget, the response correction with a one-week pilot lowers regret by 51.2% relative to a trial with the same nominal randomized rider-weeks and by 53.6% in the stricter pre-specified 10% band, and in the exact-summation analysis of these week-long requests a four-week pilot, with $1/4.5$ of the randomized rider-weeks, comes within $+0.007$ $[-0.012, +0.027]$ of an 18-week trial. In the registered primary conditions of two studies where windows, populations, rules and binding budgets change from request to request, the framework's regret is below that of a trial with the same nominal rider-weeks and below that of dose interpolation of the same pilot data at every pilot length, and reusing its one-off preparation answers 60 requests $14.1\times$ and $2.70\times$ faster than re-running the pipeline for each, with identical answers. When the paid window or a selected subpopulation changes behaviour, a window-aware correction lowers one-week regret by 0.195 against direct reuse. Against a nine-offer trial fitted with the framework's own dose curve, one-week regret is 0.055 lower; 73.5% of the gain over per-offer fits of that trial comes from sharing the tilt across offers. Public retail data and three payment rules confirm the transfer with identical exact outputs.

1 INTRODUCTION

On-demand meal delivery relies on crowdsourced couriers, or *riders*, who choose when to work and how many orders to accept (Savelsbergh & Ulmer, 2022; 2024). Demand, by contrast, peaks at meal times, in bad weather and on holidays. When too few riders are active, orders go unaccepted and

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.

are cancelled; one large platform reports several hundred thousand such cancellations a day (Wu et al., 2022). Platforms therefore add budgeted *incentive activities* to per-order pay. A typical activity reads "complete 40 deliveries this week for an extra reward; complete 55 for more", possibly with a per-completion rate that steps up at each tier or a condition on the number of active days. Tiers follow recent completions, here the weekly-volume quantiles of each rider layer. Threshold incentives change behaviour only while the threshold is still ahead (Liu et al., 2023), and effort is shaped by income targets (Chen et al., 2022; Allon et al., 2023): a tier the rider would reach anyway pays for no extra work, and a tier far out of reach is ignored. A well-placed tier sits just above the rider's usual volume, so that a little extra effort earns a clearly stated reward, to serve customers faster when service is scarce. Planning places these tiers for thousands of riders under a budget.

Operators express this task as a request. They name a rider population, an activity window that may be any subset of the coming days, an activity rule and a budget band, and they need a plan days before the window opens, within minutes. A week of n days has 2^n windows, making separate models impractical and composable forecasts necessary; payments are step functions of completions, attendance and thresholds, and rules differ between activities; and the campaigns that matter most fall on holidays and days of severe weather, where history is thinnest. Placing tiers also requires the response of completions to each candidate offer, and the clean way to measure it, a randomized controlled trial (RCT), is slow. An offer's effect is observed only after its window has closed and settled, and each group–offer cell needs many rider-weeks: nine offers for five rider groups, with 1,024 randomized observations per cell at 512 riders per group and week, occupy 18 weeks, longer than the campaign the trial should inform. Experimental data are scarce while no-offer history is abundant, a tension well known in causal inference (Kallus et al., 2018; Athey et al., 2020; 2026).

We present a request-driven framework built from four stages and seven replaceable modules, shown in Figure 1. *Prediction* by modules M1–M3 turns pre-offer histories into per-rider, per-day completion laws from which any window is composed. *Reduction* by M4 groups riders into a few activity layers, each receiving one plan, which reduces a portfolio search over individuals to a small multiple-choice problem. *Integration* by M5 and M6 corrects the no-offer law for the offered incentive and compiles joint trajectories into award probabilities: every rule, however complex, reduces to probabilities of nested award events and marked moments, from which expected payment and uplift follow, as Section 3 shows. *Allocation* by M7 selects one plan per layer inside the budget band and returns executable rewards. Every element of the request is a hyperparameter of a downstream stage, so a new period, population, rule or budget reuses every upstream output whose inputs are unchanged. The correction in M5 is designed to shorten the long trial: abundant no-offer history supplies the shape of each group's trajectory law, a pilot of one to four weeks at three offers supplies a few offer-specific moments, and history reweighted to match those moments stands in for every offer.

Contributions. (i) A general framework whose four stages exchange fixed intermediates—per rider-day laws, activity layers, award probabilities and per-layer plan menus—so that a new period, population, activity rule or budget is answered by changing hyperparameters while the response law it relies on stays valid, and each stage can be replaced and priced on its own; Sections 2–3 describe it. (ii) A history-anchored response correction from a short pilot, tested on controlled laws against a trial with the same nominal randomized rider-weeks and an 18-week trial, and in two registered studies in which windows, populations, rules and binding budgets change from request to request, one with a new rider-level generator whose response depends on the rule, and against stronger controls, including the trial fitted with the framework's own dose curve, and behaviour-changing windows and subpopulations, all in Section 5. (iii) A four-term loss decomposition—synthetic data, moment matching, integration and decision—with a conditional end-to-end perturbation bound on a fixed plan menu, constants that cannot be improved from the final tables, lower bounds attained by the statistical chain itself, and, for each stage, instances on which omitting it leaves an error floor that the other stages cannot remove; with the ablations, this shows that every stage of the combination is needed, as Section 3 sets out. (iv) An evaluation in Sections 4–6 against the practices a platform would run instead, reporting the accuracy given up, the time saved and the accuracy on cities left out of training.

2 Requests and the four-stage framework

Request. At issue time t_0 a request $r = (\mathcal{I}, W, x, \Gamma, \mathcal{B})$ names an eligible population $\mathcal{I}$, an activity window W of forecast days, context x, an activity rule Γ and a budget set $\mathcal{B}$; context covers meal

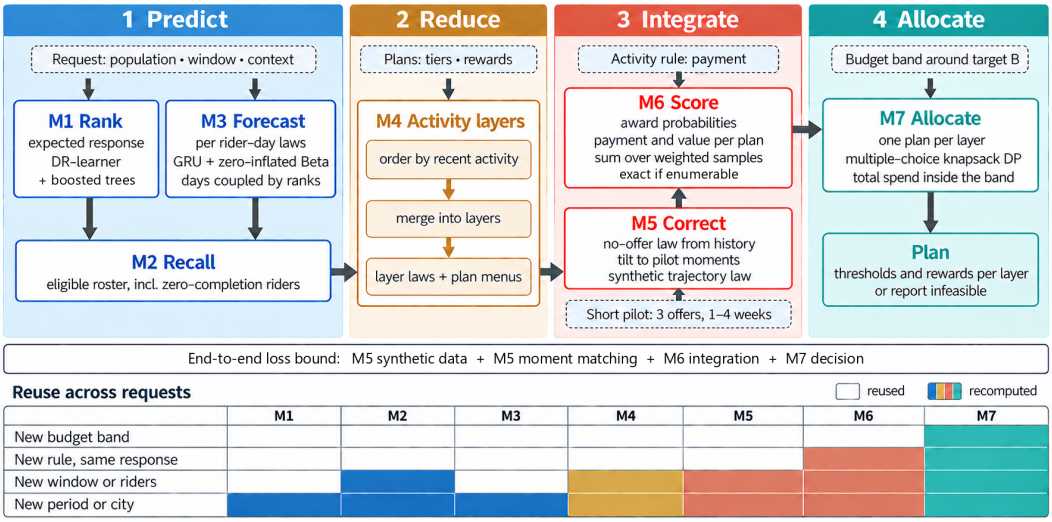

Figure 1: Request-driven framework: four stages and seven modules. Dashed chips mark inputs: request fields, candidate plans, activity rule, budget band and short pilot; arrows show data flow. The prediction stage, M1–M3, gives every eligible rider a conditional completion law; the reduction stage, M4, merges riders into activity layers with plan menus; the integration stage, M5 and M6, tilts the no-offer law to short-pilot moments and scores every plan by a weighted sum over stored samples of the corrected law, an exact sum when trajectory types can be enumerated; the allocation stage, M7, chooses one plan per layer inside the two-sided budget band and returns the rewards that attain it. Boxes name the modules used on the delivery platform; the replay of Section 4 uses gradient-boosted quantiles. DR: doubly robust; GRU: gated recurrent unit; DP: dynamic programming. The bar names the four terms of the bound of Proposition 1; the table marks what a new request recomputes.

periods, weather, supply and demand. A partition $\mathcal{G} = \{G_1, \dots, G_L\}$ of $\mathcal{I}$ assigns plan $a_l \in \mathcal{A}_l$ to group l; a plan specifies ordered thresholds, rewards or rates, eligible orders and attendance conditions. Let $Z_i(a)$ be rider i's activity trajectory in W under plan a, $g_i(Z_i(a), a)$ the rule's payment, and $V_i(a)$ an outcome such as completed deliveries. With H_{t_0} the pre-offer history,

$$c_l(a) = \sum_{i \in G_l} \mathbb{E}[g_i(Z_i(a), a) \mid H_{t_0}, x], \qquad u_l(a) = \sum_{i \in G_l} \mathbb{E}[V_i(a) - V_i(0) \mid H_{t_0}, x], \tag{1}$$

where plan 0 is the no-offer baseline. The planner solves the multiple-choice problem

$$\max_{a_l \in \mathcal{A}_l} \sum_l \widehat{v}_l(a_l) \quad \text{subject to} \quad \sum_l \widehat{c}_l(a_l) \in \mathcal{B}, \tag{2}$$

with $v_l = u_l$ for pure uplift, or v_l adding a monetary score for the probability that group spend stays within $\pm 20\%$ of its prediction. The band is part of the request: operations tolerate $\mathcal{B} = [0.8B, 1.2B]$, a stricter $[0.9B, 1.1B]$ keeps spend closer to the budget, and a sole cap $[0, B]$ is a special case.

Why four stages. Each stage removes one source of hardness. Per rider-day laws remove the 2^n windows: one forecast composes any window. Layers remove the individual search: L layers with $|\mathcal{A}_l|$ plans each leave $\prod_l |\mathcal{A}_l|$ portfolios instead of one plan per rider. Award probabilities remove the dependence on the rule: every rule is integrated from the same trajectories. Budget allocation removes the coupling between layers: a single program over the band replaces a search over portfolios. The price of each removal is an approximation error. Section 3 shows that the errors of the estimation and allocation stages add at the allocation interface on a fixed menu; the price of reduction, restricting riders to layer menus, is measured rather than bounded.

Four stages, seven modules. The framework maps a request through fixed intermediates, with the steps listed in Algorithm 1 of Appendix J,

$$(H_{t_0}, r) \xrightarrow{\text{M1–M3}} \widehat{P}_{i,d} \xrightarrow{\text{M4}} \mathcal{G}, \{\mathcal{A}_l\} \xrightarrow{\text{M5–M6}} \{\widehat{c}_l(a), \widehat{v}_l(a)\} \xrightarrow{\text{M7}} \widehat{a}. \tag{3}$$

Prediction. M1 ranks riders by expected response, M2 selects the eligible roster and M3 forecasts a completion distribution for every rider and day. Day-level laws are coupled across days by reordering each rider's forecast days according to the ranks observed in the rider's own history, so that any window W is composed from one forecast pass. *Reduction.* M4 orders riders by recent activity, merges them into contiguous activity layers and shrinks each menu only in ways that cannot remove a feasible optimum: under a sole cap it drops a plan when another costs no more and is worth no less; under a band with a floor only when another plan of the same cost is worth no less, since a cheaper plan may fall below the floor, as Lemma 1 states. *Integration.* M5 turns the no-offer law of each layer into an offer-conditioned trajectory law, and M6 integrates the rule over it as a weighted sum over stored trajectories, which is the exact sum when the trajectory types can be enumerated,

$$(\widehat{c}_l(a), \widehat{v}_l(a)) = \sum_s \bar{w}_s \big(g_l(Z_l^{(s)}(a), a), h_l(Z_l^{(s)}(a), a)\big), \quad \bar{w}_s \geq 0, \quad \sum_s \bar{w}_s = 1, \tag{4}$$

with $\bar{w}_s$ the law's probability of type s, or the normalized law weight of stored joint sample s when types cannot be enumerated. *Allocation.* M7 solves Eq. (2) by exact enumeration or a dynamic program over a budget grid, a multiple-choice knapsack (Sinha & Zoltners, 1979); along affine reward paths it runs over piecewise-affine score segments and returns the objective and reward backpointers of dense exact allocation, as Appendix J.1 shows. Returned rewards attain the reported payment.

Synthetic trajectories from a short pilot. Let $p_l^0(z)$ be the no-offer law of layer l over trajectory types z, estimated from history, and $\phi(z)$ low-dimensional trajectory features, namely normalized completions and active days. For each piloted offer a the correction is the minimum relative-entropy law that reproduces the pilot moments $\bar{\phi}_{l,a}$,

$$q_{l,a}(z) \;\propto\; p_l^0(z) \exp\big(\lambda_{l,a}^\top \phi(z)\big), \qquad \mathbb{E}_{q_{l,a}}[\phi] = \bar{\phi}_{l,a}, \tag{5}$$

an exponential tilt solved by Newton's method on the convex dual (Csiszár, 1975). Offers between pilot points interpolate λ piecewise linearly in the reward increment, with $\lambda = 0$ at the no-offer plan. Integration then sums the rule over $q_{l,a}$ by Eq. (4), with no redraw; redrawing S trajectories from $q_{l,a}$ only adds Monte Carlo error and is an ablation. History fixes the shape of the law and the pilot fixes how the offer moves it, so a pilot of one to four weeks at three offers is meant to stand in for a trial over every offer; Section 5 measures how far it does.

Generality. A new window selects other day coordinates of the same forecast; a new population re-runs M2–M4 on stored laws; a new budget re-runs M7 only; a new rule changes the payment function in M6 and, when it changes behaviour, the pilot moments in M5. Reusing stored laws is exact for computation; reusing a response correction assumes the response is unchanged, so a new city, season or rule that may change behaviour calls for its own pilot. Any activity whose payment is a function of the trajectory enters through Eq. (4), so fixed prizes, per-completion rates, attendance awards and their combinations share one pipeline. Group-additive allocation assumes no response interference between groups; joint chance constraints would need an additional state in M7.

3 Award probabilities and the four-term loss decomposition

One intermediate for every rule. For one rider and plan, let $E_1 \supseteq \cdots \supseteq E_J$ be nested award events, including attendance and eligibility conditions, with award probabilities $p_j = \Pr(E_j \mid a, H_{t_0}, x)$ and $p_{J+1} = 0$. If only the highest achieved fixed prize R_j is paid,

$$\mathbb{E}[g] = \sum_{j=1}^{J} R_j (p_j - p_{j+1}); \tag{6}$$

if the activity pays rate ρ_j on every eligible completion at the highest achieved tier,

$$\mathbb{E}[g] = \sum_{j=1}^{J} (\rho_j - \rho_{j-1}) \mathbb{E}[Y \mathbf{1}\{E_j\}], \qquad \rho_0 = 0, \tag{7}$$

with Y the eligible completions. Award probabilities and award-marked moments therefore carry everything a rule needs; the rule selects which of them are integrated. Expected spend is additive over riders without independence assumptions, whereas its variance and the probability of staying in a band depend on the joint law, which is why M6 integrates joint trajectories, not marginals: rules with equal expected payment and award probability can differ in budget risk, as Appendix J shows.

Four-term loss decomposition. For every layer l and plan a the framework passes through a chain of trajectory laws: the true offer law P; the best synthetic law P^{syn}, which applies Eq. (5) to the true no-offer law with population moments at the piloted offers and the framework's own interpolation of λ elsewhere, so that the interpolation bias at unpiloted offers belongs to this term and does not shrink with the pilot; the moment-matched law P^{mm}, which uses the estimated no-offer law and the pilot moments; and the integrated law P^{int}, which is P^{mm} under exact summation and a finite-sample law otherwise, from weighted stored samples or S redrawn trajectories in the sampled ablation. Law k induces tables $(c_l^{(k)}, v_l^{(k)})$, and allocation finally works with tables $(c^{\mathrm{dec}}, v^{\mathrm{dec}})$ on its budget grid, which equal $(c^{\mathrm{int}}, v^{\mathrm{int}})$ under exact enumeration. For $k \in \{\mathrm{syn}, \mathrm{mm}, \mathrm{int}, \mathrm{dec}\}$ with predecessor k^- in this chain, let $\varepsilon_k^c = \sum_l \max_{a \in \mathcal{A}_l} |c_l^{(k)}(a) - c_l^{(k^-)}(a)|$, define ε_k^v in the same way, and put $E_c = \sum_k \varepsilon_k^c$ and $E_v = \sum_k \varepsilon_k^v$. With $\mathcal{B}_\delta = [\underline{b} + \delta, \bar{b} - \delta]$, let $V^\star(\delta)$ be the best true value of portfolios with true expected spend in $\mathcal{B}_\delta$, and $\omega(\delta) = V^\star(0) - V^\star(\delta)$ the value of the outer margin δ.

Proposition 1. Four-term decomposition on a fixed menu. *Suppose some portfolio has true expected spend in $\mathcal{B}_{2E_c}$, and the planner returns $\widehat{a}$ with $c^{\mathrm{dec}}(\widehat{a}) \in \mathcal{B}_{E_c}$ and $v^{\mathrm{dec}}(\widehat{a}) \geq \max\{v^{\mathrm{dec}}(a) : c^{\mathrm{dec}}(a) \in \mathcal{B}_{E_c}\} - \eta$. (i)* Upper bound. *The true expected spend of $\widehat{a}$ lies in $\mathcal{B}$, and*

$$V^\star(0) - v(\widehat{a}) \;\leq\; \omega(2E_c) + 2\big(\varepsilon_{\mathrm{syn}}^v + \varepsilon_{\mathrm{mm}}^v + \varepsilon_{\mathrm{int}}^v + \varepsilon_{\mathrm{dec}}^v\big) + \eta. \tag{8}$$

On a finite menu $V^\star$ is a step function, so the relevant control is a discrete margin: $\omega(2E_c) = 0$ when an optimal portfolio has true spend in $\mathcal{B}_{2E_c}$, and if $\omega(\delta) \leq \kappa\delta$ on $[0, 2E_c]$ the right side is at most $\sum_k 2(\kappa\varepsilon_k^c + \varepsilon_k^v) + \eta$. (ii) Lower bound for tables. *For any nonnegative stage errors there are table instances on a fixed menu with exactly these errors on which Eq.* (8) *holds with equality and on which every deterministic planner that sees the final tables and keeps the true spend in $\mathcal{B}$ for all truths consistent with them loses at least $\omega(2E_c) + 2E_v$. Without the tightening, the true spend can leave $\mathcal{B}$ by E_c. (iii)* Error floors of omitted stages. *There are instances on which omitting one stage leaves an error that exact other stages cannot reduce: without the pilot correction all uplift is invisible; two matched moments without the history shape leave the probability of reaching a threshold t standard deviations above the mean undetermined within $1/(1+t^2)$; plug-in means in place of integration misprice a threshold prize by half; and an equal budget split can forfeit all value.*

The proposition is a conditional perturbation bound, given the stage errors and the tightened band, an error analysis rather than a certificate computable at planning time; with the truth known it applies to 0.5%, 2.8% and 10.0% of feasible requests at 1, 2 and 4 pilot weeks and improves on the trivial bound in 0.3%, 1.2% and 5.8%, as Table 10 reports. It is relative to the layer menus: the value given up by restricting individuals to layer menus enters only through $V^\star$; Table 2 compares alternative layerings on rider records. The lower bound shows that no planner seeing only the final tables can improve the constants. For the chain of laws itself, the minimum relative-entropy tilt of Eq. (5) attains the constant of the synthetic term exactly, and the pilot and sampling terms are attained in order but not in constant, by Proposition 3.

Why the combination is accurate and fast. Costs and pure-uplift values are expectations, hence linear in the trajectory law, so errors of the four stages add at the allocation interface instead of compounding: a value error costs at most twice its size and a cost error at most 2κ times its size under the growth control $\omega(\delta) \leq \kappa\delta$ in Eq. (8). Each term is driven by a different resource and shrinks independently: the synthetic term by the part of the response that the matched features do not explain, the moment term as $n^{-1/2}$ in pilot size for a given base, the integration term as $S^{-1/2}$ and to zero under exact summation, and the decision term by one grid step per layer with upward rounding and half a step with nearest rounding. An estimated base adds the term of Lemma 6, with proofs in Appendix K. Corollary 1 states the rates in expectation, for a stability score centred at the estimated mean under a margin condition. The stage errors belong to the compiled tables, not to a request, so every budget and weight answered from the same tables inherits them; this is why paying for the upstream stages once gives large speed-ups without adding error terms. On a given instance errors may partly cancel, so the ablations below measure what each stage buys in practice.

4 Rider records: accuracy given up and time gained

Data and protocol. The rider panel holds daily completed deliveries of 3,000 riders over 595 days in 2025–2026 across 17 separately modelled cities and one pooled node. We replay the framework at

Table 1: Accuracy given up and time gained against the most accurate way to answer a request without the framework, on the same inputs. Speed-up: median paired ratio, or of geometric-mean times for the mixed-integer program; brackets: interquartile range or 95% interval over laws.

Alternative per request	Accuracy the framework gives up	Time: alternative → framework	Speed-up
Rider records: 3,000 riders, 45 weekly origins, five activity layers			
Re-draw trajectories and solve exactly	none in the scenarios, with 1,431/1,431 requests identical; allocation loses value in 3/405 instances, at most 0.92%	117.8 s → 3.07 s per week-long request	38.6× [30.8, 43.0]; 11.0× over 127 requests with the one-off pass
One plan per rider, exact	not measurable: the exact plan did not finish within 3 hours	unfinished → 4.05 s per table	—
Controlled response laws: 18 laws × 8 replicates, 33 budgets each, Section 5			
Randomized trial over all offers	relative regret, planned and scored within 20% of budget: 0.113 → 0.121 with a 4-week pilot, difference +0.007 [−0.012, +0.027]; a trial with the pilot's rider-weeks: 0.421 with 1 week, framework 0.205, and 0.225 with 4	9,216 → 2,048 nominal randomized rider-weeks per group, 18 → 4 weeks	1/4.5 as many
Public retail data from M5 and three payment rules, Section 6			
Re-draw and solve exactly on 3,000 M5 series	none: identical scenarios, value lost in 0/405 tables	84.8 s → 4.36 s per request	20.0× [18.9, 21.2]; 7.11× over 127 with preparation
Mixed-integer program per request, model kept between requests, 396 conditions	none: the budget program is exact on 65,340/65,340 requests; the MIP on 65,315 raw, 65,335 envelope	paired on conditions both solve exactly	1, 9, 33 requests: raw 1.08, 7.6, 20.9×; envelope 1.06, 2.5, 5.5×

45 weekly origins with three sampling seeds: models are refit at each origin on data before it, and each seven-day window is planned before its outcomes are read. Plans combine three tier shapes, six active-day requirements and six reward levels, 108 plans per layer, so five activity layers give $108^5 \approx 1.5 \times 10^{10}$ portfolios; budgets are upper caps at $\beta = 0.5$, 1.0 and 1.5 times the realized cost of a reference plan, so they are set in hindsight. Because every plan is scored against the completions riders actually recorded, the replay measures what a plan must get right before any behavioural response: who attains which tier and what the plan costs. A plan's value is the number of deliveries completed by riders who reach an award, its cost the reward paid; plan-cost error is the median absolute relative error between predicted and realized plan cost over plans with nonzero realized cost, which excludes the 810 of 72,900 plans that pay nothing, and regret is measured against the exact optimum on the realized records. Recorded completions do not respond to a plan, so value does not depend on the reward level, and the budget binds in hindsight only in the Spring Festival week, in 3 of 135 units at $\beta = 0.5$. The replay therefore tests pricing and computation; Section 5 tests the choice of reward level and allocation under a binding budget. M3 fits gradient-boosted quantiles with rider shares, and intervals resample origins; Appendix A gives the details.

What the framework is compared with. We compare with the most accurate practices a platform would otherwise run, each exact or unbiased for the step it replaces: re-drawing trajectories for the requested window and solving exactly, request by request; one exact plan per rider; a mixed-integer program (MIP) per request (Sinha & Zoltners, 1979) solved to optimality (Huangfu & Hall, 2018); and a trial long enough to observe every offer. Table 1 reports accuracy given up and time saved.

Near-lossless and much faster. Against re-drawing and solving exactly per request, the framework gives up nothing in trajectories and almost nothing in allocation, while a week-long request takes 3.07 s instead of 117.8 s. Each such request follows a one-off sampling pass of 114.2 s per origin, the only cold-start cost. Charging that pass to the framework, eight windows are answered 1.81× faster and all 127 windows of a week take 12.2 instead of 131.1 minutes, 11.04× faster over the 9 timed origins of Figure 2b, fixed before the run; solving one plan per rider is out of reach, which is what

Table 2: Rider records, 45 weekly origins × 3 seeds, five activity layers. Each row replaces one stage of the full framework by a cheaper alternative. Ratios are of origin means with 95% intervals over origins. Regret at budget fraction 0.5; pp: percentage points. *45 origins × 3 budget fractions.

Stage	Full framework	Cheaper alternative	Accuracy change
Prediction	per rider-day distribution	point forecast	plan-cost error ×3.08 [2.48, 3.81]
Reduction	activity layers	equal-width buckets	regret 0.37% → 0.78%
Integration	joint law across days	independent days	plan-cost error ×1.51 [1.33, 1.74]
		two-moment normal	×2.46 [2.02, 3.01]; 5–6-day award −10.31 pp
		plug-in mean	×1.85 [1.58, 2.19]
Allocation	budget-grid program	equal budget split	value −15.7% [13.8, 17.4]
		exact reference solver	loss 0 in 134/135*; max 0.69%

population reduction buys. The gain comes from reuse, which the interface allows: per-day laws are fixed before a request arrives, so one forecast serves every request, as timed in Table 5.

Every stage earns its place. Table 2 replaces one stage at a time by the cheaper practice it displaces. Distributional prediction is the largest single gain: a point forecast raises plan-cost error from 0.215 to 0.660 and median award-rate error from 4.0 to 66.9 percentage points. Joint integration is next: composing days independently multiplies plan-cost error by 1.51, and a two-moment normal baseline, which takes window totals as normal and prices tiers without the truncated-moment term, multiplies it by 2.46. Equal-width buckets change the layer menus and raise regret against the same hindsight optimum from 0.37% to 0.78%, and an equal budget split gives up 15.7% of the value found by the budget program. On the realized five-layer records the grid program equals the exact optimum in 134 of 135 main-grid instances, with a largest loss of 0.69% as Appendix J.1 details.

Cities left out of training. The replay refits every model weekly on all cities. A held-out condition fixed before the runs removes the evaluated cities from training, as Appendix A.1 details: cities removed in four folds over 45 origins and forecast from their riders' own history keep plan-cost error at 0.343 against 0.340 for a reference trained on them, a paired change of $+0.003$ $[-0.018, +0.020]$ within the pre-set margin of 0.045. Every stage replacement in Table 8 again loses accuracy on them.

5 Synthetic trajectories and a short pilot versus a long trial

Design. Behavioural response cannot be read from logged completions, so we test the response-correction step M5 on controlled response laws where the truth is known; every other stage is fixed. We draw 24 new laws in four families of six, whose responses are saturating, threshold and hump combined, saturating only, threshold only, or absent; each law has five rider groups under three payment rules, nine offers per group and 32 weekly trajectory types; allocation enumerates all 9^5 portfolios for 33 budgets, and every request covers the whole week and all five groups. Each of the eight replicates per law gives each method its own data: 4,096 no-offer weeks per group; a *long RCT* of 1,024 randomized observations per group and offer, which at 512 riders per group and week takes 18 weeks; a *same-calendar RCT* that spreads the rider-weeks of a pilot evenly over all nine offers; and the *framework*, which runs the pilot at three offers for one, two or four weeks, corrects history by Eq. (5) and, since the 32 types can be enumerated, sums the rule exactly over the corrected law, which removes the integration term by Lemma 7; drawing $S = 2{,}048$ trajectories instead is an ablation. The endpoint is relative regret of pure uplift against the true optimum in the request's band; an abstention scores zero, and so does a plan whose true expected spend leaves the band. We report the operating band $[0.8B, 1.2B]$, chosen after the runs, in which every method plans and is scored and the optimum is taken, and the pre-specified, stricter $[0.9B, 1.1B]$ alongside, which the appendix tables of this design use. Laws are the unit of inference, with 95% intervals by bootstrap over the 18 laws with a response; the 6 null laws check validity only. Arms were specified before the runs, and every conclusion also holds without the one law used to fix the operator, as Appendix B shows.

Results. Table 3 gives three findings; the pre-specified 10% band, plotted in Figure 2a of Appendix B, follows each 20% value. First, for the same nominal randomized rider-weeks the framework is far better than randomizing the pilot over all offers: regret falls by 0.215 $[-0.261, -0.173]$ with a

Table 3: Relative regret, 18 response laws × 8 replicates × 33 budgets; planning, optimum and scoring within 20% of the budget or the pre-specified 10%. Indented rows change one stage; Table 9 has the rest. Validity: % of issued plans with true expected spend in the 10% band over all 24 laws.

	Within 20% of budget			Within 10%, pre-specified		
Method	1 wk	2 wk	4 wk	1 wk	2 wk	4 wk
Long RCT, 18 weeks, business alternative		0.113			0.130	
Same-calendar RCT	0.421	0.313	0.225	0.542	0.383	0.264
Framework: tilt, exact sum	**0.205**	**0.159**	**0.121**	**0.252**	**0.186**	**0.136**
sampled integration, $S = 2{,}048$	0.245	0.197	0.163	0.286	0.223	0.189
without history: pilot frequencies	0.603	0.560	0.527	0.658	0.611	0.569
Validity, framework / same-calendar				80.3/53.2	83.5/69.0	87.5/76.5

Table 4: Two registered studies with changing requests and binding budgets: registered primary condition, 36 response laws each, relative regret within 20% of the budget. A: fresh laws from the generator of Table 3; B: a new rider-level generator with a rule-dependent response. Speed-up: 60 requests, one-off preparation charged, against re-running the pipeline per request.

	A: integrated study			B: new generator		
Pilot	1 wk	2 wk	4 wk	1 wk	2 wk	4 wk
Same-calendar RCT	0.391	0.296	0.216	0.477	0.326	0.240
Framework	**0.220**	**0.172**	**0.148**	**0.188**	**0.150**	**0.116**
Dose interpolation, same information	0.227	0.177	0.154	0.210	0.171	0.131
Independent days, least costly replacement	0.311	0.292	0.296	0.198	0.158	0.124
Budgets that bind, % of all 48 laws		72.3			75.0	
Speed-up, 60 requests	14.1×	14.2×	14.3×	2.70×	2.72×	2.78×

one-week pilot, 51.2% lower, and by 0.290 [−0.338, −0.243] at 10%, 53.6% lower; with two and four weeks it falls by 0.155 [−0.198, −0.115] and 0.105 [−0.140, −0.071], and by 0.197 and 0.129 at 10%. Second, in the exact-summation analysis of these week-long requests, a four-week pilot, with 4.5 times fewer randomized rider-weeks, comes within +0.007 [−0.012, +0.027] of the 18-week trial and within +0.006 [−0.013, +0.024] at 10%; the registered test of the sampled-integration arm is listed in Table 29. Third, under 10% scoring each stage is needed against the replacement tested for it: dropping history, moment matching or joint integration raises one-week regret by +0.406, +0.404 and +0.331, and an equal budget split on the exact tables raises it by +0.610, each interval excluding zero; drawing 2,048 trajectories instead of summing exactly raises it by +0.039 [+0.017, +0.064], and by +0.035 [+0.010, +0.061] at 10%.

Changing requests, binding budgets and a new generator. Studies A and B combine response, changing requests and binding budgets; see Table 4 and Appendices C–D. Each replicate draws 60 random windows, populations and rules, with budgets at 0.3, 0.5 and 0.7 times unconstrained optimal cost; one-off preparation is charged. Study A draws 48 fresh laws from the preceding generator. Study B uses an unrelated generator with six groups of 16 latent rider types, day-of-week activity and rule-dependent response. Its weeks cannot be enumerated, so requests use weighted sums over 4,096 stored weeks per group. In the registered primary condition of both studies, both registered hypotheses of Table 29 hold: regret is below the same-calendar trial's at every pilot length, by −0.171 [−0.201, −0.140] and −0.289 [−0.311, −0.267] with one week, and reusing the prepared laws for 60 requests is faster than re-running the pipeline for each, with identical answers. Every stage replacement and dose interpolation of the same history and pilot data, a same-information control, lose accuracy at every pilot length. In three more registered studies, detailed in Appendix E, all hypotheses hold: at matched cache the framework beats a per-offer tilt of the trial, a normal law with exact truncated moments and the trial under shared calendar-week shocks, and request-aware reuse beats direct reuse when windows or subpopulations change behaviour. The third, reported in Table 17, fits the trial jointly with the framework's own dose curve: at the same parameterization the framework's regret is lower by 0.055 [0.051, 0.059] at one week, and of its one-week advantage over the per-offer tilt, sharing the tilt across offers accounts for 73.5% [72.1, 74.8] and placing the pilot at three offers for 26.5%.

In a fourth, five history layers retain 94.6% [94.4, 94.8] of a type-informed continuous-population optimum, and with 20 riders per group 98.4% of the same-information exact integer optimum, as Appendix E reports. In the registered sweeps of Appendix E the advantage over the joint fit holds for responder shares down to 1/8, curvature 0.25 between piloted doses and, with unreachable targets projected and all fits converged, history-support gaps up to 25%. At a fixed pilot and history, three and five layers beat one plan per group from 135 to 8,640 rider-weeks per group in Appendix E.

Appendix F gives two learned routes the framework's information, cache and settlement: a cross-validated density-ratio classifier, and monotone response curves feeding the same knapsack, an instance of the estimate-then-allocate route of Albert & Goldenberg (2022); their one-week regret exceeds the framework's by 0.356 [0.347, 0.364] and 0.020 [0.018, 0.022], a relative reduction of 9.3% against the response model. The classifier takes 36.0× the framework's time. A frozen policy spends a third of the pilot on three untested offers, refits on disagreement and falls back to dose interpolation when a target is unreachable or a payment check fires. Over the ten conditions of study G it beats the equal-effort trial under every tail and support-gap departure and at curvature up to 0.5 and keeps 81.5% [80.5, 82.6] of the framework's gain under the baseline generator M0; curvature between piloted offers, invisible in the piloted cells, is flagged in 27% and 54% of units at 0.5 and 1, where regret falls by 0.069 [0.062, 0.075] and 0.278 [0.268, 0.288] against the unchecked framework.

6 Transfer across domains and rules

The same modules transfer by changing inputs and hyperparameters, with exact-allocation checks in Appendix H. On 3,000 M5 item–store series (Makridakis et al., 2022a;b), the same rule represents a supplier's tiered rebate. Reuse cuts median request time from 84.8 s to 4.36 s, a paired 20.0× speed-up in request time, with identical scenarios and exact values on all 405 tables; with preparation charged, the 127 requests of an origin are answered 7.11× faster. Across 396 conditions with fixed prizes, per-completion rates and attendance awards, reusing the budget–value curve solves all 65,340 requests exactly. Per-request mixed-integer programs leave 25 requests unresolved in raw form and 5 in envelope form, and take 20.9× and 5.5× longer at 33 requests, respectively; see Table 1.

7 Related work

Response or uplift models feed budgeted allocation in marketing (Zhao et al., 2019; Albert & Goldenberg, 2022; Ai et al., 2022; Zhou et al., 2024; Sun et al., 2024; Zhang et al., 2025; Yang et al., 2026; Cong et al., 2025); delivery bonuses and driver or rider subsidies are allocated alike (Wu et al., 2022; Chen et al., 2024; Yang et al., 2025; Chen et al., 2026; 2025). Decision-focused learning combines observational and experimental data to train the allocation model itself (Zhang et al., 2025); here the pilot enters through a few moments of a trajectory law, so one correction serves every rule and window. The decision unit is a multi-day threshold activity paid on the whole trajectory, and requests change window, population, rule and budget. Our allocation stage extends multiple-choice knapsacks and graphical Bellman algorithms (Sinha & Zoltners, 1979; Kameshwaran & Narahari, 2009; Gafarov et al., 2014) to two-sided budget bands and reward traceback. Where small experiments complement observational data to estimate effects (Kallus et al., 2018; Athey et al., 2020; 2026), our pilot corrects history by entropy balancing (Hainmueller, 2012) on offer-conditioned trajectories coupled across days by rank (Clark et al., 2004; Schefzik et al., 2013), and its error is its own term of the end-to-end bound.

8 Conclusion

Four stages reuse conditional trajectory laws across changing requests for multi-day threshold activities. The four-term bound prices each stage on a fixed menu, as an error analysis rather than a certificate computed at planning time, and reuse answers all weekly windows 11.04× faster at almost no loss. Exact summation over a given law and exact allocation on a given grid do not make the response law exact; the controlled studies measure that part, and the framework improves on matched-information trials and learned controls within the tested ranges, by 0.020 [0.018, 0.022] in one-week regret against the closest learned route. Rider records without randomized offers support pricing and computation; response evidence covers the tested generators, menus and conditions.

AI USE STATEMENT

In this work, we used generative AI tools to edit the paper to improve readability and to create or edit software code. Supporting qualitative and thematic data analysis, formulating questions for surveys or interviews and transcribing recordings are not applicable to this work. We have reviewed all AI-assisted work: AI-assisted code was tested before use, and every reported number is computed from recorded experimental outputs by analysis scripts. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The rider replay uses historical records of riders who took part in voluntary incentive activities, whose tiers are set so that a little extra effort earns a clearly stated reward. The paper reports only aggregates of these records; no individual rider record is reported or released, and the supplementary material contains no rider records. The controlled studies use synthetic data drawn from closed-form generators and involve no human subjects. Because budgeted incentives decide which riders receive which offers, eligibility fairness and rider welfare deserve evaluation alongside efficiency whenever such a framework is deployed.

REPRODUCIBILITY STATEMENT

Algorithm 1 lists the steps that answer a request. Appendix A specifies the rider replay; Appendices B, C, D, E and F specify the synthetic-trajectory experiment and the registered studies A–K with their endpoints and decision rules; Appendices G, H and I specify the public and controlled computations and Appendix J the settlement interface and exact allocation; Appendix K states the assumptions and gives complete proofs. The anonymous supplementary material contains the generators, the correction, integration and allocation code, the registered designs and analysis scripts, the per-request and per-law results and analysis outputs of the registered studies; for the public retail data, the series list and panel construction, the forecasting, sampling, plan-replay and timing code with its configuration and a driver that re-runs one origin, the request list, the settlement and allocation code and the per-origin timing and exact-value results; for the rider replay, per-origin timing and allocation aggregates without rider records; and a script that checks Tables 1, 10, 17, 20, 21, 22, 23, 24, 25, 26 and 27, the public-data results of Appendix H and the corresponding rows of Tables 29, 30, 31 and 32 against them and checks the exact and budget-grid allocation solvers against brute force. Appendix L lists every registered test family.

REFERENCES

Meng Ai, Biao Li, Heyang Gong, Qingwei Yu, Shengjie Xue, Yuan Zhang, Yunzhou Zhang, and Peng Jiang. LBCF: A large-scale budget-constrained causal forest algorithm. In *Proceedings of the ACM Web Conference 2022 (WWW '22)*, pp. 2310–2319, 2022. doi: 10.1145/3485447.3512103.

Javier Albert and Dmitri Goldenberg. E-commerce promotions personalization via online multiple-choice knapsack with uplift modeling. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM '22)*, pp. 2863–2872, 2022. doi: 10.1145/3511808.3557100. arXiv:2108.13298.

Gad Allon, Maxime C. Cohen, and Wichinpong Park Sinchaisri. The impact of behavioral and economic drivers on gig economy workers. *Manufacturing & Service Operations Management*, 25 (4):1376–1393, 2023. doi: 10.1287/msom.2023.1191.

Susan Athey, Raj Chetty, and Guido Imbens. Using experiments to correct for selection in observational studies. *arXiv preprint arXiv:2006.09676*, 2020. Revised May 2025.

Susan Athey, Raj Chetty, Guido W. Imbens, and Hyunseung Kang. The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. *The Review of Economic Studies*, 93(4):2284–2312, 2026. doi: 10.1093/restud/rdaf087. Published online 30 September 2025.

Bobby Chen, Siyu Chen, Jason Dowlatabadi, Yu Xuan Hong, Vinayak Iyer, Uday Mantripragada, Rishabh Narang, Apoorv Pandey, Zijun Qin, Abrar Sheikh, Hongtao Sun, Jiaqi Sun, Matthew Walker, Kaichen Wei, Chen Xu, Jingnan Yang, Allen T. Zhang, and Guoqing Zhang. Practical marketplace optimization at Uber using causally-informed machine learning. In *2nd Workshop on Causal Inference and Machine Learning in Practice, KDD 2024*, 2024. arXiv:2407.19078.

Juhua Chen, Karson Shi, Jialing He, North Chen, and Kele Jiang. MMCE: A framework for deep monotonic modeling of multiple causal effects. *arXiv preprint arXiv:2504.03753*, 2025.

Taijie Chen, Rui Su, Siyuan Feng, Laoming Zhang, Hongyang Zhang, Haijiao Wang, Zhaofeng Ma, Jintao Ke, and Li Ma. D^3-Subsidy: Online and sequential driver subsidy decision-making for large-scale ride-hailing market. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '26)*, pp. 7094–7105, 2026. doi: 10.1145/3770855.3818350.

Xirong Chen, Zheng Li, Liu Ming, and Weiming Zhu. The incentive game under target effects in ridesharing: A structural econometric analysis. *Manufacturing & Service Operations Management*, 24(2):972–992, 2022. doi: 10.1287/msom.2021.1002.

Martyn Clark, Subhrendu Gangopadhyay, Lauren Hay, Balaji Rajagopalan, and Robert Wilby. The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 5(1):243–262, 2004. doi: 10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2.

Yu Cong, Chao Xu, and Yi Zhou. Large-scale trade-off curve computation for incentive allocation with cardinality and matroid constraints. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)*, pp. 2575–2582, 2025. doi: 10.24963/ijcai.2025/287.

Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975. doi: 10.1214/aop/1176996454.

Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning (ICML)*, pp. 1097–1104, 2011.

Evgeny R. Gafarov, Alexandre Dolgui, Alexander A. Lazarev, and Frank Werner. A graphical approach to solve an investment optimization problem. *Journal of Mathematical Modelling and Algorithms in Operations Research*, 13(4):597–614, 2014. doi: 10.1007/s10852-013-9248-2.

Evgeny R. Gafarov, Alexandre Dolgui, Alexander A. Lazarev, and Frank Werner. A new effective dynamic program for an investment optimization problem. *Automation and Remote Control*, 77(9):1633–1648, 2016. doi: 10.1134/S0005117916090101.

Jens Hainmueller. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46, 2012.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Q. Huangfu and J. A. J. Hall. Parallelizing the dual revised simplex method. *Mathematical Programming Computation*, 10(1):119–142, 2018. doi: 10.1007/s12532-017-0130-5.

Nathan Kallus, Aahlad Manas Puli, and Uri Shalit. Removing hidden confounding by experimental grounding. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/566f0ea4f6c2e947f36795c8f58ba901-Abstract.html.

S. Kameshwaran and Y. Narahari. Nonconvex piecewise linear knapsack problems. *European Journal of Operational Research*, 192(1):56–68, 2009. doi: 10.1016/j.ejor.2007.08.044.

Tianming Liu, Zhengtian Xu, Daniel Vignon, Yafeng Yin, Qingyang Li, and Zhiwei Qin. Effects of threshold-based incentives on drivers' labor supply behavior. *Transportation Research Part C: Emerging Technologies*, 152:104140, 2023. doi: 10.1016/j.trc.2023.104140.

Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4):1346–1364, 2022a. doi: 10.1016/j.ijforecast.2021.11.013.

Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos, Zhi Chen, Anil Gaba, Ilia Tsetlin, and Robert L. Winkler. The M5 uncertainty competition: Results, findings and conclusions. *International Journal of Forecasting*, 38(4):1365–1385, 2022b. doi: 10.1016/j.ijforecast.2021.10.009.

Martin W. P. Savelsbergh and Marlin W. Ulmer. Challenges and opportunities in crowdsourced delivery planning and operations. *4OR*, 20(1):1–21, 2022. doi: 10.1007/s10288-021-00500-2.

Martin W. P. Savelsbergh and Marlin W. Ulmer. Challenges and opportunities in crowdsourced delivery planning and operations—an update. *Annals of Operations Research*, 343(2):639–661, 2024. doi: 10.1007/s10479-024-06249-1.

Roman Schefzik, Thordis L. Thorarinsdottir, and Tilmann Gneiting. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, 28(4):616–640, 2013.

Prabhakant Sinha and Andris A. Zoltners. The multiple-choice knapsack problem. *Operations Research*, 27(3):503–515, 1979. doi: 10.1287/opre.27.3.503.

Zexu Sun, Hao Yang, Dugang Liu, Yunpeng Weng, Xing Tang, and Xiuqiang He. End-to-end cost-effective incentive recommendation under budget constraint with uplift modeling. In *Proceedings of the 18th ACM Conference on Recommender Systems (RecSys '24)*, pp. 560–569, 2024. doi: 10.1145/3640457.3688147. arXiv:2408.11623.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. doi: 10.1007/b13794.

Zhuolin Wu, Li Wang, Fangsheng Huang, Linjun Zhou, Yu Song, Chengpeng Ye, Pengyu Nie, Hao Ren, Jinghua Hao, Renqing He, and Zhizhao Sun. A framework for multi-stage bonus allocation in meal delivery platform. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*, pp. 4195–4203, 2022. doi: 10.1145/3534678.3539202. arXiv:2202.10695.

Jiaqi Yang, Lexiao Chen, Zicheng Su, Wanjing Ma, Zhichao Zou, and Kun An. Decision-focused learning for optimal subsidy allocation in ride-hailing services. *Transportation Research Part C: Emerging Technologies*, 180:105301, 2025. doi: 10.1016/j.trc.2025.105301.

Siyun Yang, Shixiao Yang, Jian Wang, Di Fan, Kehe Cai, Haoyan Fu, Jiaming Zhang, Wenjin Wu, and Peng Jiang. Jointly optimizing debiased CTR and uplift for coupons marketing: A unified causal framework. *arXiv preprint arXiv:2602.12972*, 2026. Accepted at CIKM 2026.

Shuli Zhang, Hao Zhou, Jiaqi Zheng, Guibin Jiang, Bing Cheng, Wei Lin, and Guihai Chen. Bi-level decision-focused causal learning for large-scale marketing optimization: Bridging observational and experimental data. In *Advances in Neural Information Processing Systems 38 (NeurIPS 2025)*, 2025. arXiv:2510.19517.

Kui Zhao, Junhao Hua, Ling Yan, Qi Zhang, Huan Xu, and Cheng Yang. A unified framework for marketing budget allocation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*, pp. 1820–1830, 2019. doi: 10.1145/3292500.3330700. arXiv:1902.01128.

Hao Zhou, Rongxiao Huang, Shaoming Li, Guibin Jiang, Jiaqi Zheng, Bing Cheng, and Wei Lin. Decision focused causal learning for direct counterfactual marketing optimization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*, pp. 6368–6379, 2024. doi: 10.1145/3637528.3672353. arXiv:2407.13664.

A RIDER REPLAY: PROTOCOL AND TIMING

Panel and calendar. The panel holds daily completed deliveries of 3,000 riders over a 595-day observation span, recorded per rider, city and day; a missing rider-day means no recorded completion. Seventeen cities with enough riders are modelled as separate nodes and the rest are pooled. No feature uses information from after the origin. Forecast origins are 45 consecutive weeks; every model is refit at each origin on data up to the origin, and a request window is any subset of the seven days that follow.

Stages as run. The replay instance of M3 combines quantile gradient-boosted forecasts of node totals with per-rider share and concentration models into per rider-day completion distributions, and Table 28 lists a recurrent network as another instance of M3; R joint samples per origin and seed are coupled across days by reordering each rider's forecast days according to the ranks observed in the rider's own 182-day history. Reduction forms five activity layers, or three coarser ones, from participation in the preceding 28 days. Integration averages each plan's payment and value over the R joint samples. Allocation solves the multiple-choice problem on a 400-point budget grid with costs rounded up, so every portfolio it finds respects the cap; the exact reference enumerates each layer's menu after cost–value dominance, which is exact here because every budget is a sole cap, by Lemma 1. The main-grid instances in which the grid program loses value fall in the Spring Festival week.

Plans, value and budget. A layer's plan is a ladder of three rungs, an active-day requirement $d \in \{0, 2, 3, 4, 5, 6\}$ and a reward level $\rho \in \{0.25, 0.5, 0.75, 1, 1.5, 2\}$, 108 plans in all; a coarse grid keeps 18 and a day-requirement grid keeps $d \in \{5, 6\}$. The rungs are the 50/70/85%, 60/80/92% or 40/65/85% quantiles of the layer's seven-day totals over the 182 days before the origin. With Y_i the rider's window completions, D_i its active days and J_i the highest rung reached when $D_i \geq d$ and $J_i = 0$ otherwise, the plan pays $\rho\, m_{J_i} Y_i$ with $m = (0, 1, 1.6, 2.4)$. The value is the award-covered volume $\sum_i \mathbf{1}\{J_i \geq 1\} Y_i$. It is a pricing target, not the uplift of Eq. (1): with recorded completions held fixed, ρ changes only the cost. No stability weight enters the objective. The budget is a cap $B = \beta$ times the realized cost of a reference portfolio with the middle ladder, $d = 3$ and $\rho = 1$ in every layer; because this cost is read from the evaluated week, the budget is retrospective, whereas operationally a request fixes B in advance.

Endpoints and uncertainty. Plan-cost error is $|\widehat{c} - c|/c$ for predicted and realized plan cost, summarized by its median over the plans of a unit; the 810 plans with $c = 0$, 1.1% of 72,900, are excluded and counted, all in the dormant layer. Award-rate error is the predicted minus realized share of riders with $J_i \geq 1$, over all plans. Regret is $(V^\star - V(\widehat{a}))/V^\star$, with $V(\widehat{a})$ the realized value of the portfolio chosen on the predicted table and $V^\star$ the exact optimum of the realized table of the full framework's layering at the same budget, found by enumeration over each layer's cost–value Pareto set without rounding any cost, while the portfolio $\widehat{a}$ itself is chosen by the budget-grid program; $V^\star > 0$ in every unit. Because every replacement is graded against this one optimum, regret can be negative: a portfolio that overspends can exceed $V^\star$, and a replacement with other layers has other menus. The framework's regret is nonzero only in the two weeks around the Spring Festival, 0.37% on average at $\beta = 0.5$. Equal split gives every layer B/L and takes, per layer, the plan of largest predicted value whose predicted cost fits the share. The replacements are a point forecast, the rounded per rider-day mean; equal-width buckets of the 28-day active-day count; independent days, the same per-day draws without coupling; a two-moment normal shortcut, which takes a per-rider normal window total, prices each tier's payment and covered volume as the mean times the normal tier probability without the truncated-mean correction, and judges the day requirement on mean active days; and a plug-in mean. Each origin is first averaged over its three seeds; 95% intervals use an origin-block bootstrap with blocks of one and two origins and 20,000 draws, reporting the wider interval. Every origin, including the four Spring Festival weeks, is kept.

Timing. All timings in Table 5 are single-threaded, at 45 origins, seed 1. A framework request reduces the stored per-day draws to the requested days, integrates the rule over the 540 plans and allocates. The alternative re-draws the per-day samples of the requested days with the same random streams, re-couples them and runs the same integration and exact allocation; the re-drawn samples equal the stored ones exactly at every request, so the comparison is of time only. Model fitting is

needed by both sides once per origin and is shown separately. The first request takes 2.97 s against 3.07 s for later ones, so the only cold-start cost is the 114.2 s sampling pass. $K = 8$ uses all 45 origins; $K = 32$ and 127 use origins 0, 5, 10, 15, 20, 25, 30, 35, 40, fixed before the run. The complete per-request alternative of Table 1, 117.8 s, also recomputes the rank template; re-drawing alone takes 114.5 s. By window length, the framework and the alternative take 2.39 and 18.5 s for one day and 5.63 and 101.1 s for six days.

Table 5: Where the time of rider requests goes, in seconds, medians over origins, single-threaded. Top: one week-long request, by step. Bottom: K requests per origin in the fixed order of whole week, seven single days and remaining subsets; the framework total includes its once-per-origin sampling pass; speed-ups are medians of per-origin ratios. [a]Gradient-boosted node quantiles, 29.8 s, plus share and concentration models, 189.1 s, needed equally by both approaches. [b]Speed-up when model fitting is charged once to both sides.

Step	Framework	Re-draw per request
Model fitting, once per origin[a]	221.3 once	221.3 once
Per-day draws, 7 days	108.7 once	108.7
Cross-day rank template	3.5 once	reused
Cross-day coupling	1.6 once	1.8
Window reduction	0.54	0.54
Integration, 540 plans	2.52	2.52
Allocation	0.44 ms, grid	0.50 ms, exact
Per request, once-per-origin steps excluded	3.07	114.5
Same, first request	2.97	—

K	Origins	First	Total	Per req.	Re-draw	Ratio	+Fitting[b]
8	45	117.8	136.1	17.01	245.0	1.81	1.31
32	9	117.7	240.4	7.51	1410.7	5.87	3.58
127	9	117.7	734.5	5.78	7864.2	11.04	8.82

Time saved by cheaper shortcuts. Table 6 adds, for the shortcuts of Table 2, the time each would save. Drawing and coupling trajectories run once per origin; window reduction, integration, which takes 2.52 of the 3.07 s of a week-long request, and allocation run for every request.

Table 6: Cheaper shortcuts the framework declines, on rider records, five layers, main plan grid. Accuracy: plan-cost error of the shortcut over the framework's, or value lost, with 95% intervals over origins. Time: the framework's stage time over the shortcut's.

Stage	Shortcut	Accuracy cost of the shortcut	Time the shortcut saves
Prediction	point forecast	plan-cost error $\times 3.08$ [2.48, 3.81]	not timed separately
Integration	independent days	plan-cost error $\times 1.51$ [1.33, 1.74]	1.4% of sampling; integration $\times 1.001$
Integration	two-moment normal	plan-cost error $\times 2.46$ [2.02, 3.01]; 5–6-day plans $\times 3.29$	integration 85.8$\times$ of 3.86 s per table
Allocation	equal budget split	value -15.7% [13.8, 17.4]	0.51 $\rightarrow$ 0.01 ms

A.1 Cities left out of training

The reference refits the forecasting models at every weekly origin on all cities. For the held-out condition the 17 separately modelled city nodes are ranked by median daily participation and assigned to four folds of 5, 4, 4 and 4 nodes; the pooled node of small cities always stays in training. For each fold and each of the 45 origins, all training rows of the fold's cities, and the all-city aggregate rows, are removed, together with the fold's contexts in the share and concentration models, and the fold's riders are forecast from their own history only. All later stages, menus and budgets are unchanged, and the seed is that of the reference, so both use the same random streams. At each

origin the held-out result is paired with the reference on the same riders and outcomes; intervals resample origins with $B = 20{,}000$ draws, averaging folds within an origin. The pre-set rule calls a loss immaterial if the upper 95% bound of the paired increase in plan-cost error at budget fraction 1 is at most 0.045, the half-width of the reference's own interval, and that of selection regret at most 0.01. All 180 runs finished; nothing is dropped, and a control run that reuses the reference models at one origin reproduces the reference window exactly. Table 7 gives the result, and Table 8 the stage replacements on the held-out riders. The design fixed before the runs also contained a condition with models fitted once and left unrefit; every pre-specified rule and its outcome is listed in Table 29.

Table 7: Cities left out of training, rider records, five-layer system, budget fraction 1. Plan-cost error is the median absolute relative error of predicted plan cost; change = held-out minus reference on the same riders and outcomes, 95% interval over origins. Pre-set margin: upper bound ≤ 0.045.

Condition	Origins / runs	Riders per run	Plan-cost error Held-out	Plan-cost error Reference	Change	Within margin
Unseen cities, four folds	45 / 180	311	0.343	0.340	+0.003 [−0.018, +0.020]	yes

Table 8: Stage replacements on held-out riders: replacement minus full framework, budget fraction 1, 95% interval over origins. Prediction and integration are scored by plan-cost error, reduction and allocation by selection regret. Unseen cities: the condition of Table 7; fit once: models fitted at the start of a 16-week evaluation span and used throughout that span.

Replacement	Endpoint	Fit once	Unseen cities
Point forecast	plan-cost error	+0.416 [+0.359, +0.472]	+0.380 [+0.292, +0.467]
Plug-in integration	plan-cost error	+0.317 [+0.235, +0.382]	+0.123 [+0.097, +0.148]
Two-moment normal	plan-cost error	+0.459 [+0.382, +0.523]	+0.205 [+0.167, +0.240]
Independent days	plan-cost error	+0.202 [+0.142, +0.250]	+0.058 [+0.035, +0.080]
Equal-width buckets	regret	+0.0017 [+0.0006, +0.0028]	+0.0046 [+0.0024, +0.0070]
Equal budget split	regret	+0.048 [+0.021, +0.080]	+0.122 [+0.099, +0.145]

B SYNTHETIC TRAJECTORIES AND A SHORT PILOT: DESIGN AND FURTHER RESULTS

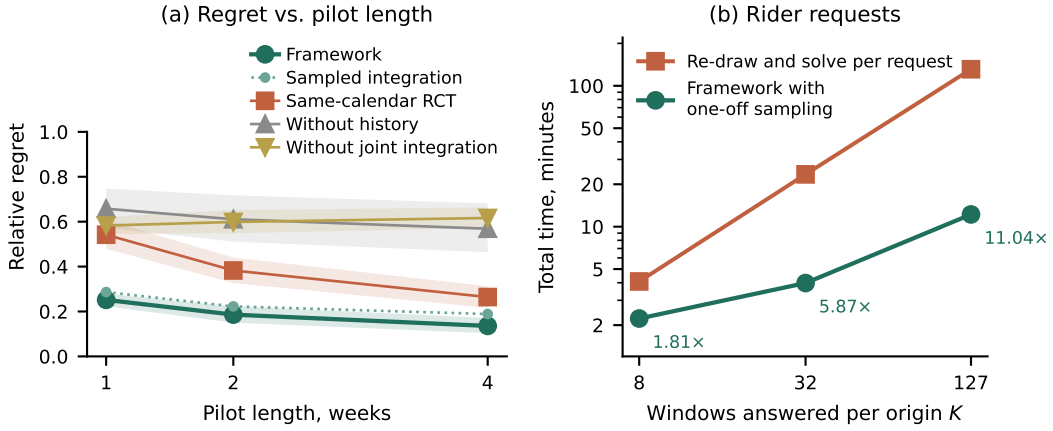

Figure 2: (a) Relative regret against pilot length, 18 response laws, 10% scoring; bands: 95% intervals over laws. (b) Total time to answer K windows per origin on rider records, one-off sampling pass charged; labels: median paired speed-ups.

Laws. The 24 laws use the support, offers, payment rules and budgets of Appendix I with their own response generator and a fresh seed: five groups, 32 weekly trajectory types, nine offers per

group whose reward-rate increments form a 3×3 grid, and group rules fixed award, per-completion, attendance, fixed award and per-completion. With dose $x = 2b + t$ of offer $a = 3b + t$, completions C_k and active days D_k of type k, base weights $w_{gk} \in \{1, \dots, 17\}$ and a weekly shock $z \in \{0, 1\}$ that equals one with probability $\frac{1}{2}$, is drawn independently for every sample index and is shared by all groups and offers at that index, the no-offer law is proportional to $w_{gk} s_{gk}(z)$ with

$$s_{gk}(z) = 64 + u_g[zC_k + (1 - z)(C_{\max} - C_k)],$$

and the law under dose x to

$$w_{gk} s_{gk}(z) \left[256 + a_g \min(x, \tau_g) C_k + b_g (x - \tau_g)_+^2 D_k + c_g x(6 - x)(C_{\max} - C_k)\right],$$

with integers $a_g \in \{1, \dots, 4\}$, $b_g \in \{1, \dots, 6\}$, $c_g \in \{0, \dots, 3\}$, $\tau_g \in \{1, \dots, 4\}$ and $u_g \in \{1, 2, 3\}$. The combined family keeps all three response terms; the saturating family sets $b_g = c_g = 0$, the threshold family $a_g = c_g = 0$ and the null family $a_g = b_g = c_g = 0$. The response form is unknown to every method.

Methods. All methods share exact enumeration over 9^5 portfolios, the same 33 budgets $B \in \{0, 5, \dots, 160\}$ and the band $[0.9B, 1.1B]$. The long RCT uses offer-conditional frequencies from 1,024 observations per group and offer. The same-calendar RCT uses the pilot's rider-weeks spread uniformly over the nine offers. Budgets are nominal: with integer division over offers, the nominal 512, 1,024 and 2,048 rider-weeks per group give 510, 1,023 and 2,046 pilot observations and 504, 1,017 and 2,043 same-calendar observations. The framework pilots offers with increments 0.2, 0.6 and 1.0, solves Eq. (5) exactly for two moments, normalized completions and active days, at each piloted offer, interpolates the tilt linearly in the increment through zero at the no-offer plan, and sums payment and value exactly over the 32 types. Ablations change one stage each: *sampled integration* averages 2,048 draws per group and offer from the same corrected law; *without history* uses the pilot frequencies directly, with unpiloted offers taking the nearest piloted offer, an arm called "without synthesis" in the pre-specified rules; *without moment matching* uses the history law for every offer; *without joint integration* evaluates payment and value on the mean trajectory; *without allocation* splits the budget equally and lets each group choose its best offer within its share; the pre-specified version, listed in Table 29, ran on the sampled tables, and Section 5 and Table 9 apply it to the exact-sum tables, a re-analysis of the stored laws without new draws; within 20% of the budget it raises one-week regret by $+0.455$ [$+0.344$, $+0.571$].

Endpoint. For one replicate and pilot length, relative regret is $\sum_r (U_r^\star - U_r) / \sum_r U_r^\star$ over the budgets r whose true optimum is feasible, with $U_r^\star$ the optimal true uplift and U_r the true uplift of the issued plan, set to zero for an abstention or a plan whose true expected spend leaves the band. Budgets without a feasible portfolio enter only the validity count; a replicate whose optimal uplifts sum to zero, as in the null laws, has no regret. Replicates are averaged within a law.

Pre-specified design. Design, endpoints, arms, unit of inference and decision rules were fixed before the runs. The operator, one tilt per piloted offer at three offers, was fixed after one run of the pipeline on one law with one replicate and a one-week pilot, in place of a single tilt with a linear dose; that law remains in the analysis. The pre-specified primary arm integrated 2,048 sampled trajectories; the exact sum over the 32 types was pre-specified as the finite-pilot arm of the stage decomposition, and we use it as the default because it removes the integration term by Lemma 7; the pre-specified rules and their outcomes for the sampled arm are listed in Table 29. Every conclusion also holds on the other 17 response laws: the framework minus the same-calendar RCT is -0.296 [-0.345, -0.248], -0.198 [-0.243, -0.157] and -0.135 [-0.169, -0.101] at one, two and four weeks; the gap to the long RCT at four weeks is $+0.004$ [-0.015, $+0.023$]; and the ablations raise one-week regret by $+0.386$, $+0.391$, $+0.328$ and, for sampling, $+0.034$ [$+0.008$, $+0.061$].

The decomposition, measured. Because the truth is known, each term of Eq. (8) can be computed; Table 10 does so for the sampled variant, which carries the synthetic, moment-matching and integration terms; allocation enumerates exactly, so the decision term is zero, and exact summation also sets the integration term to zero. As a separate allocation diagnostic, the budget-grid program is within 0.0015 relative regret of exact enumeration. Lengthening the pilot shrinks only the moment-matching

term: each doubling divides it by 1.40 and 1.41, the $\sqrt{2} \approx 1.41$ rate of Corollary 1(a) for uplift; the synthetic term, which includes the interpolation bias, does not depend on the pilot and is the fixed price of shortening experimentation. As a heuristic, the decomposition suggests lengthening a pilot until the moment term is comparable to the synthetic term, which happens at two weeks here; realized regret keeps falling with a longer pilot, as Table 3 shows. Proposition 1 needs the true cost error, so it is an after-the-fact error analysis; on the 26, 140 and 499 requests with one-, two- and four-week pilots whose comparator fits in the band tightened by twice that error it holds without exception.

Table 9: Further stage replacements of the framework: penalized relative regret, pre-specified 10% scoring, 18 response laws. The first two arms were not rescored at 20%; the last row gives the equal split on the exact tables within 20% of the budget. Their differences from the framework are in Section 5.

Replacement	1 wk	2 wk	4 wk
Without moment matching: history only	0.656	0.656	0.656
Without joint integration: plug-in mean	0.583	0.599	0.616
Without budget allocation: equal split, exact tables	0.861	0.860	0.855
the same, scored within 20% of the budget	0.661	0.656	0.647

B.1 Empirical check of the decomposition

In panel A each stage term is measured against the preceding law, exactly as in the proposition, and the direct error is below the sum of stage terms, as the triangle inequality allows. The synthetic and integration terms do not change with the pilot, while the moment term falls by factors 1.40 and 1.41 per doubling of the pilot, and by 1.44 and 1.44 for cost, close to the $\sqrt{2} \approx 1.41$ of Corollary 1(a). Part (i) held on every request where its comparator exists. The check tightens by the direct error $\bar{E}_c$, which needs the truth, so it is an after-the-fact error analysis. Among the 4,968 requests whose band holds a portfolio of true cost, the tightened band still holds one in 26, 140 and 499, that is 0.5%, 2.8% and 10.0% at one, two and four weeks, and the bound is below the comparator value, the trivial bound, in 16, 62 and 290, or 0.3%, 1.2% and 5.8%; a rule held to the tightened band would abstain on the other 4,942, 4,828 and 4,469; realized regrets are reported in Section 5. The grid program is indistinguishable from enumeration. In panel B the signs follow the mechanisms of Appendix K: composing days independently understates the 5–6-day requirements and overstates lower ones, the crossing expected under shared daily shocks; the normal law with attendance evaluated at its mean fails where attendance binds, the plug-in floor of part (iii); and dropping distributional prediction or joint allocation costs most. Panel B concerns quantities that can be replayed under the offered plans; it involves no behavioural response.

C Integrated study: changing requests and binding budgets

Design, endpoints and decision rules were registered, with the simulation code, before the run. The analysis script was completed after the run and pooled all 48 laws for timing; H2 is reported here, as registered, on the 36 response laws.

Laws and requests. The 48 laws are fresh draws of the generator of Appendix B, twelve per family; 36 have a response, and the 12 null laws check validity and timing only. Behaviour responds to the offer at the week level, and the request decides which days are paid and counted. Every replicate issues its own stream of 60 requests. Each request draws, uniformly, an activity window among the 127 non-empty subsets of the seven days, a population among the 26 subsets of at least two of the five groups, and one rule for all its groups, which is a fixed award, a per-completion rate or attendance. Tier thresholds scale with the window, $\max(1, \mathrm{round}(t\,|W|/7))$ for $t \in \{14, 28, 42\}$ completions or $\{2, 4, 6\}$ active days; uplift is the change in completions inside the window. With c_{ref} the true expected cost of the cheapest portfolio that maximizes uplift without a budget, or the median portfolio cost when the largest uplift is zero, the request carries three budgets $B = \beta c_{\mathrm{ref}}$, $\beta \in \{0.3, 0.5, 0.7\}$, which gives 180 budgeted requests per replicate; the budgets are set from the true law so that they bind, and every method receives the same ones. A budget binds when its band optimum is below the unconstrained maximum; over all 48 laws, 72.3% of budgeted requests bind.

Table 10: Empirical check of Proposition 1. **A**: controlled response laws with known truth, 24 laws $\times$ 8 replicates, 33 budgets each, sampled variant with $S = 2{,}048$, which carries the synthetic, moment-matching and integration terms; under exact summation the integration term is zero and the other terms are unchanged; allocation enumerates exactly, so $\varepsilon_{\mathrm{dec}} = 0$, and the last two rows of A are an allocation diagnostic comparing the budget-grid program with enumeration; value errors are uplift in completions and cost errors are currency units, both summed over five layers; the median optimal uplift of a request is 14.1. **B**: rider records, 45 weekly origins $\times$ 3 seeds, five activity layers; each row replaces one stage of the framework and is compared with the framework on the realized records; intervals resample origins.

A Pilot length	1 week	2 weeks	4 weeks
Synthetic term $\varepsilon^v_{\mathrm{syn}}$ / $\varepsilon^c_{\mathrm{syn}}$	5.92 / 5.53	5.92 / 5.53	5.92 / 5.53
Moment term $\varepsilon^v_{\mathrm{mm}}$ / $\varepsilon^c_{\mathrm{mm}}$	8.82 / 7.67	6.29 / 5.31	4.48 / 3.69
Integration term $\varepsilon^v_{\mathrm{int}}$ / $\varepsilon^c_{\mathrm{int}}$, $S = 2{,}048$	3.43 / 3.00	3.44 / 3.00	3.51 / 3.00
Sum of stage terms	18.16 / 16.20	15.65 / 13.83	13.90 / 12.22
Direct error $\bar{E}_v$ / $\bar{E}_c$, truth vs. final tables	11.79 / 10.88	10.06 / 9.35	8.89 / 8.24
Requests with a comparator in $\mathcal{B}_{2\bar{E}_c}$	26/6336	140/6336	499/6336
share of the 4,968 feasible requests	0.5%	2.8%	10.0%
bound below the comparator value	16	62	290
Violations of part (i), tightening by $\bar{E}_c$	0	0	0
Grid program − enumeration, relative regret	0.0005	0.0015	0.0009
95% interval over laws	$[-0.0014, 0.0028]$	$[-0.0005, 0.0038]$	$[-0.0014, 0.0033]$

B Stage replaced by	Mechanism	Observed on rider records
Prediction: point forecast	plug-in floor, part (iii)	plan-cost error $\times 3.08$ [2.48, 3.81]; median award-rate error 4.0 $\rightarrow$ 66.9 pp
Integration: independent days	cross-day dependence	award rate -3.07 pp $[-3.28, -2.87]$ on 5–6-day plans, $+3.32$ pp [2.93, 3.69] on the others†; cost error $\times 1.51$
Integration: per-rider normal, attendance at its mean	two moments; plug-in gate	-10.31 pp $[-10.97, -9.69]$ on 5–6-day plans, $+0.23$ pp $[-0.19, 0.57]$ on the others†; cost error $\times 2.46$
Integration: plug-in mean	plug-in floor, part (iii)	$+10.17$ pp [9.31, 11.03] on the others†, -4.11 pp on 5–6-day plans; cost error $\times 1.85$
Decision: budget grid vs. exact optimum	Lemma 8	zero loss in 134 of 135 five-layer main-grid instances, 45 origins $\times$ 3 budgets, max 0.69% on realized-record tables
Decision: equal budget split	allocation floor, part (iii)	value -15.7% [13.8, 17.4] at budget fraction 0.5

†Plans requiring 0, 2, 3 or 4 active days, derived exactly from the all-plan and 5–6-day means. Award-rate entries are differences from the framework's joint law in percentage points.

Data and methods. Each of eight replicates per law draws 4,096 no-offer weeks per group, a pilot of one, two or four weeks at 512 riders per group and week spread over offers 2, 5 and 8, and a same-calendar RCT that spreads the same nominal rider-weeks over all nine offers; with integer division per cell this gives 510, 1,023 and 2,046 pilot and 504, 1,017 and 2,043 same-calendar observations per group for one, two and four weeks. In the registered primary condition reported here, every sample index has its own weekly shock shared by all cells. The *framework* prepares its laws once per origin, from history and pilot only: the history law, the moment-matched tilt of Eq. (5) at the piloted offers interpolated in the reward increment, and the corrected law of every offer. Each request then reuses these laws: its tables come from the exact sum over the 32 trajectory types for the request's window, population and rule, and its plan from exact enumeration over the $9^{|\text{population}|}$ portfolios in the band. The same-calendar RCT uses its empirical offer laws with the same integration and allocation. Stage replacements change one stage: without moment matching uses the history law for every offer; without history uses pilot frequencies at the nearest piloted offer; plug-in integration pays at the mean window completions and active days; independent days composes days from per-day marginals; equal split divides the band equally over the population's groups. We also report the budget-grid program with 1,024 states in place of enumeration and, as a same-information

control, dose interpolation, which interpolates the history and pilot laws in the reward increment. For timing, every request is also answered by re-running the whole pipeline, including the preparation, from the raw data; its answers must equal the framework's.

Endpoints and decision rules. Relative regret is computed per replicate over the budgeted requests whose band optimum exists, scoring an abstention or a plan whose true expected spend leaves the band as zero; replicates are averaged within a law and 95% intervals come from 20,000 bootstrap resamples of the 36 response laws. The primary band is $[0.8B, 1.2B]$, with $[0.9B, 1.1B]$ alongside. The confirmatory family of Table 29 has two hypotheses. H1: the framework has lower regret than the same-calendar RCT at each pilot length, with the upper bound below zero. H2: answering all 60 requests of an origin by reuse with the preparation charged is faster than re-running the pipeline per request, with the lower bound of the law-mean log speed-up over the 36 response laws above zero and every timed answer identical. The breakdowns of Table 12 are descriptive; a subgroup is summarized over the laws in which it contains a request with a feasible band optimum, and at $\beta = 0.3$ two laws have none; this summary rule was fixed after the run.

Table 11: Integrated study: relative regret, lower is better, on 36 response laws × 8 replicates × 180 budgeted requests, with each request's window, population, rule and budget drawn afresh. Indented rows change one stage of the framework; every difference from the framework has a 95% interval above zero. Validity: share of the framework's issued plans whose true expected spend lies in the band, over all 48 laws.

	Within 20% of budget			Within 10%		
Method	1 wk	2 wk	4 wk	1 wk	2 wk	4 wk
Same-calendar RCT	0.391	0.296	0.216	0.527	0.388	0.269
Framework	**0.220**	**0.172**	**0.148**	**0.258**	**0.198**	**0.164**
without moment matching	0.592	0.592	0.592	0.626	0.626	0.626
without history	0.471	0.433	0.404	0.561	0.509	0.470
plug-in integration	0.592	0.604	0.631	0.669	0.672	0.690
independent days	0.311	0.292	0.296	0.368	0.340	0.339
equal split	0.699	0.692	0.684	0.902	0.898	0.892
Dose interpolation, same information	0.227	0.177	0.154	0.265	0.205	0.169
Validity of the framework, %	93.2	94.0	94.1	87.7	90.0	91.1

Results. Framework minus same-calendar RCT is −0.171 [−0.201, −0.140], −0.124 [−0.160, −0.089] and −0.069 [−0.104, −0.035] within 20%, that is 43.8%, 42.0% and 31.8% lower, and −0.269 [−0.297, −0.240], −0.190 [−0.225, −0.156] and −0.105 [−0.139, −0.073] within 10%. With a one-week pilot, the stage replacements raise regret by +0.373 [+0.333, +0.412] without moment matching, +0.251 [+0.191, +0.312] without history, +0.372 [+0.328, +0.417] with plug-in integration, +0.091 [+0.068, +0.116] with independent days and +0.479 [+0.392, +0.570] with an equal split. The budget-grid program's regret differs from that of enumeration by at most 0.001 in every cell, and dose interpolation has higher regret by +0.007 [+0.004, +0.010], +0.005 [+0.003, +0.007] and +0.006 [+0.003, +0.009] at one, two and four weeks. Table 12 shows that the gain over the same-calendar RCT holds in every rule, window length, population size and budget level, and Table 13 gives the preparation-charged timing.

D Response validation with a new generator

This study answers whether the results of Appendix C depend on the generator that produced its laws. Its design, endpoints, decision rules, simulation code and analysis script were all registered before the run.

Generator. None of it is shared with the 32-type construction of Appendix B. Each of 48 laws, twelve per family, has six rider groups, and each group is a mixture of 16 latent rider types with Dirichlet weights.

Table 12: Integrated study by request type, within 20% of budget, descriptive: framework regret and framework minus same-calendar RCT with 95% intervals over the laws in which the subgroup is defined.

			1-week pilot		2-week pilot	
	Level	Laws	Framework	minus same-calendar	Framework	minus same-calendar
Rule	fixed award	36	0.225	−0.175 [−0.208, −0.142]	0.176	−0.127 [−0.166, −0.089]
	per completion	36	0.229	−0.172 [−0.206, −0.138]	0.180	−0.125 [−0.165, −0.087]
	attendance	36	0.205	−0.165 [−0.192, −0.139]	0.162	−0.120 [−0.151, −0.089]
Window days	1–2	36	0.221	−0.193 [−0.226, −0.159]	0.173	−0.146 [−0.184, −0.110]
	3–5	36	0.219	−0.171 [−0.200, −0.141]	0.171	−0.123 [−0.159, −0.088]
	6–7	36	0.226	−0.155 [−0.194, −0.113]	0.172	−0.119 [−0.159, −0.082]
Groups	2	36	0.216	−0.214 [−0.238, −0.190]	0.171	−0.152 [−0.183, −0.123]
	3	36	0.215	−0.160 [−0.193, −0.126]	0.169	−0.116 [−0.154, −0.076]
	4	36	0.230	−0.153 [−0.190, −0.114]	0.183	−0.111 [−0.154, −0.067]
	5	36	0.223	−0.148 [−0.187, −0.111]	0.186	−0.105 [−0.148, −0.063]
Budget β	0.3	34	0.305	−0.290 [−0.346, −0.237]	0.255	−0.205 [−0.253, −0.160]
	0.5	36	0.228	−0.197 [−0.258, −0.151]	0.180	−0.151 [−0.216, −0.101]
	0.7	36	0.176	−0.149 [−0.181, −0.118]	0.131	−0.109 [−0.145, −0.075]

Table 13: Integrated study, preparation charged, in milliseconds, single-threaded; medians are over replicates for the one-off preparation and over requests otherwise. Speed-up: time to answer the first K requests of a replicate by re-running the whole pipeline for each, over the preparation plus the reusing requests; geometric mean over the 36 response laws, with replicates averaged on the log scale, and 95% intervals. Answers are identical on all 51,840 timed requests of these laws and on all 69,120 of all 48 laws.

Pilot	Preparation	Request, reuse	Request, re-run	$K = 8$	$K = 32$	$K = 60$
1 week	7.27	0.22	7.57	5.8×	11.9×	14.1× [13.8, 14.5]
2 weeks	7.40	0.22	7.57	5.8×	11.9×	14.2× [13.8, 14.5]
4 weeks	6.25	0.21	6.61	5.9×	12.0×	14.3× [14.0, 14.7]

- *Activity.* A type is active on day d with probability $\mathrm{logit}^{-1}(\alpha_k + \gamma_d)$, where γ_d is a day-of-week effect of the group.
- *Completions.* An active day has $1 + \mathrm{Poisson}(\lambda_k - 1)$ completions, truncated at 15, and a weekly shock multiplies λ_k by $1 \pm \delta$.
- *Response.* The response depends on the rule shown. Under dose x of rule ρ, activity becomes $p + (1 - p)\min\{0.9, A_k S^A_\rho f(x)\}$ and intensity $\lambda(1 + B_k S^B_\rho f(x))$. The sensitivities are $(S^A, S^B) = (0.4, 0.8)$ for a fixed award, $(0.2, 1)$ for a per-completion rate and $(1, 0.2)$ for attendance.
- *Families.* Both channels with a saturating f; intensity only; a threshold f in which only types whose usual week lies between the first and third tier respond fully; no response.

Weeks take 16^7 daily patterns per rider, so the framework cannot enumerate them. The truth is exact: a mixture over types and shocks of day-by-day convolutions of the law of window completions and active days.

Requests, data and methods. Requests follow Appendix C, with populations of two to four of the six groups; over all 48 laws, 75.0% of budgeted requests bind. Each of eight replicates per law draws:

- 4,096 no-offer weeks per group;
- a pilot of one, two or four weeks at 512 riders per group and week, spread over the three rules times offers 2, 5 and 8;
- a same-calendar RCT that spreads the same nominal rider-weeks over the three rules times all nine offers. With integer division per cell the actual observations per group are 504, 1,017 and 2,043 in the pilot and 486, 999 and 2,025 in the same-calendar RCT for one, two and four weeks, so the trial has slightly fewer.

For every group and rule the *framework* tilts the stored weeks to the pilot moments by Eq. (5), interpolated in the reward increment. It does this once per origin; every request then reuses the weights and integrates its window rule as a weighted sum over the stored weeks, as in lines 5–6 of Algorithm 1. The stage replacements are the same as in Appendix C: history weeks for every offer; pilot weeks at the nearest piloted offer; payment at the weighted mean; days composed independently; equal split. Dose interpolation mixes history and pilot weeks with interpolation weights in the reward increment. For timing, each request re-fits the tilts it needs from the raw data. The endpoints, the unit of inference, here 36 response laws, and H1–H2 are as in Appendix C.

Table 14: New generator: relative regret, lower is better, on 36 response laws × 8 replicates × 180 budgeted requests. Indented rows change one stage of the framework; every difference from the framework, and that of dose interpolation, has a 95% interval above zero. Validity: share of the framework's issued plans whose true expected spend lies in the band, over all 48 laws.

	Within 20% of budget			Within 10%		
Method	1 wk	2 wk	4 wk	1 wk	2 wk	4 wk
Same-calendar RCT	0.477	0.326	0.240	0.765	0.601	0.415
Framework	**0.188**	**0.150**	**0.116**	**0.270**	**0.197**	**0.140**
without moment matching	0.433	0.433	0.433	0.781	0.781	0.781
without history	0.415	0.354	0.309	0.610	0.517	0.448
plug-in integration	0.447	0.479	0.510	0.576	0.577	0.591
independent days	0.198	0.158	0.124	0.314	0.222	0.162
equal split	0.498	0.477	0.466	0.793	0.775	0.760
Dose interpolation, same information	0.210	0.171	0.131	0.305	0.223	0.159
Validity of the framework, %	93.1	94.6	95.7	83.5	88.9	92.7

Results.

- *Same-calendar RCT.* Framework minus same-calendar RCT is −0.289 [−0.311, −0.267], −0.176 [−0.192, −0.160] and −0.123 [−0.140, −0.108] within 20%, that is 60.6%, 53.9% and 51.4% lower, and −0.495 [−0.516, −0.473], −0.404 [−0.426, −0.383] and −0.276 [−0.296, −0.257] within 10%. The gain holds in every rule, window length, population size and budget level, as Table 15 shows.
- *Stage replacements.* With a one-week pilot they raise regret by:
  - +0.245 [+0.214, +0.275] without moment matching;
  - +0.228 [+0.206, +0.252] without history;
  - +0.260 [+0.233, +0.287] with plug-in integration;
  - +0.011 [+0.009, +0.013] with independent days;
  - +0.311 [+0.269, +0.355] with an equal split.
- *Dose interpolation.* Its regret is higher by +0.022 [+0.018, +0.027], +0.020 [+0.016, +0.024] and +0.015 [+0.010, +0.019] at one, two and four weeks.
- *Allocation.* The budget-grid program's regret is within 0.001 of that of enumeration.
- *Timing.* Here the per-request alternative re-fits only the tilts the request needs. With the preparation charged, answering the first 8, 32 and 60 requests of an origin is 1.22×, 2.33× and 2.70× [2.67, 2.73] faster with a one-week pilot, and 60 requests are 2.72× and 2.78× faster at two and four weeks. Medians are 52.4 ms for the preparation, 4.1 ms per reusing request and 13.7 ms per re-fitting request. Answers are identical on all 51,840 timed requests of the response laws.

E Stronger controls, weight diagnostics, departures from the tilt family, behaviour-changing requests and population reduction

Seven further studies were registered, with designs, decision rules, simulation code and analysis scripts fixed before the runs. All use the rider-level generator of Appendix D with fresh seeds, 60

Table 15: New generator by request type, within 20% of budget, descriptive: framework regret and framework minus same-calendar RCT with 95% intervals over laws.

			1-week pilot		4-week pilot	
	Level	Laws	Framework	minus same-calendar	Framework	minus same-calendar
Rule	fixed award	36	0.179	−0.318 [−0.343, −0.293]	0.120	−0.127 [−0.143, −0.112]
	per completion	36	0.196	−0.322 [−0.343, −0.300]	0.111	−0.127 [−0.144, −0.111]
	attendance	36	0.202	−0.183 [−0.213, −0.155]	0.132	−0.114 [−0.137, −0.093]
Window days	1–2	36	0.186	−0.340 [−0.365, −0.315]	0.119	−0.154 [−0.175, −0.133]
	3–5	36	0.188	−0.285 [−0.308, −0.263]	0.116	−0.121 [−0.138, −0.106]
	6–7	36	0.189	−0.269 [−0.294, −0.242]	0.122	−0.111 [−0.130, −0.092]
Groups	2	36	0.211	−0.301 [−0.318, −0.285]	0.131	−0.154 [−0.168, −0.140]
	3	36	0.188	−0.294 [−0.315, −0.273]	0.117	−0.121 [−0.139, −0.104]
	4	36	0.176	−0.277 [−0.306, −0.248]	0.110	−0.111 [−0.131, −0.093]
Budget β	0.3	36	0.261	−0.275 [−0.300, −0.252]	0.159	−0.182 [−0.217, −0.150]
	0.5	36	0.181	−0.288 [−0.312, −0.264]	0.114	−0.118 [−0.138, −0.100]
	0.7	36	0.170	−0.301 [−0.326, −0.276]	0.106	−0.106 [−0.121, −0.093]

changing requests per replicate, budgets at 0.3, 0.5 and 0.7 times the unconstrained optimum, and the law as the unit of inference, with 95% intervals from a law bootstrap and Holm correction within each family; studies C–E and G use eight replicates per law. Regret is relative regret of pure uplift within 20% of the budget, as in Table 4.

Study C: stronger same-information controls at matched cache. It draws 384 laws, 96 per family, of which 288 have a response. The pilot and the same-calendar trial have the same number of observations per group: 486, 999 and 2,025 at one, two and four weeks under the independent condition, and 486, 972 and 1,944 under calendar-block shocks. Every arm reads the same cached 4,096 stored weeks per group and the same data, and every arm is solved by the same exact enumeration.

- *History-anchored tilt of the trial.* The framework's tilt is fitted at each of the eight non-zero offers to that offer's cell of the same-calendar trial. It uses the same history and the same number of randomized observations as the framework; study E adds a joint fit of the trial.
- *Normal law with exact truncated moments.* It keeps the framework's weights and replaces only the law of the paid statistic by a normal with the weighted mean and variance. Every tier is priced by

$$\mathbb{E}[X\mathbf{1}\{\ell \le X < h\}] = \mu\{\bar{\Phi}(z_\ell) - \bar{\Phi}(z_h)\} + \sigma\{\varphi(z_\ell) - \varphi(z_h)\},$$

  with $z = (t - 0.5 - \mu)/\sigma$ at every integer threshold t. This separates distribution shape from the moment computation.
- *Calendar condition.* A rider keeps its type through all observed weeks. One shock per group and calendar week is shared by every rider observed in that week, and the pilot, the trial and 512 concurrent no-offer riders share the pilot weeks' shocks. The framework then anchors its moment targets to the concurrent riders, using the pilot mean minus the concurrent mean plus the history mean, and the trial measures uplift against the same concurrent riders.

Results of study C. All seven registered hypotheses of Table 30 hold.

History anchoring helps the trial as well: the tilted trial has lower regret than the raw trial by 0.075 [0.072, 0.078] at one week.

The framework's regret is below that of every same-information control at every pilot length, as Table 16 shows. At one week it is below dose interpolation by 0.022 [0.021, 0.024], below the normal law with exact truncated moments by 0.053 [0.051, 0.055], and below the tilted trial by 0.209 [0.201, 0.217]. Study E below splits this last difference into its two sources. The normal comparison isolates the distribution shape, since the weights and the tier-moment formula are exact in both arms.

Under calendar shocks shared by the riders of a week, the concurrently anchored framework has lower regret than the concurrently controlled trial by 0.267 [0.259, 0.274] at one week and 0.214 [0.206, 0.222] at two weeks.

Table 16: Study C: relative regret within 20% of the budget, 288 response laws × 8 replicates × 60 requests × 3 budgets. Every arm uses the same stored weeks, data and exact enumeration. Calendar: repeated riders with one shock per group and calendar week.

Condition	Method	1 wk	2 wk	4 wk
Independent	**Framework**	**0.196**	**0.154**	**0.126**
	Dose interpolation, same information	0.219	0.171	0.138
	Normal law, exact truncated moments	0.250	0.208	0.179
	History-anchored tilt of the trial	0.405	0.291	0.211
	Same-calendar RCT	0.480	0.340	0.241
Calendar	**Framework**, concurrent anchor	**0.223**	**0.215**	**0.207**
	Same-calendar RCT, concurrent control	0.490	0.429	0.386

Reuse gives identical answers in all 184,320 comparisons per condition and pilot length. Charging the one-off preparation, it answers the 60 requests of an origin 2.73× [2.71, 2.74] faster than re-running the pipeline per request at one week, and 2.71× and 2.73× faster at two and four weeks.

Study E: placement of the observations versus sharing across offers. The tilted trial differs from the framework in two ways: its randomized observations are spread over nine offers instead of three, and its tilt is fitted separately at every offer instead of along one dose curve. Study E adds a *joint fit of the trial* that keeps the trial's data and removes the second difference. It uses the framework's own parameterization, $\lambda(s)$ piecewise linear in the reward increment s with $\lambda(0) = 0$ and nodes $\theta = (\theta_2, \theta_5, \theta_8)$ at the three piloted offers, and fits θ to all eight non-zero trial cells by maximum likelihood of the tilted-history model:

$$\widehat{\theta} = \arg\max_{\theta} \sum_{a=1}^{8} n_a \Big[\lambda_\theta(s_a)^\top m_a - \log \sum_{i=1}^{N} \exp\big\{ \lambda_\theta(s_a)^\top \phi_i \big\} \Big],$$

where m_a is the mean feature vector of the n_a trial observations at offer a and ϕ_i are the features of the N stored weeks. The problem is concave, and when the data sit at the three nodes only it separates into the framework's per-node moment matching. The difference between the framework and the tilted trial then splits into two contrasts,

$$R_{\mathrm{FW}} - R_{\mathrm{tilt}} = \underbrace{(R_{\mathrm{FW}} - R_{\mathrm{joint}})}_{\text{placement}} + \underbrace{(R_{\mathrm{joint}} - R_{\mathrm{tilt}})}_{\text{sharing across offers}},$$

the first at the same parameterization and the second on the same data. The study reuses the 384 laws of study C with fresh requests and data, the same observation counts, stored weeks, cache and exact enumeration. The joint fit is the maximum-likelihood fit, computed by a Newton method whose step is scaled as a whole and accepted only when a line search does not decrease the log-likelihood, as fixed by an amendment registered before the reported run; the registered check that it reproduces the framework's per-node fit on pilot data holds in every replicate, with a largest deviation of 2.1×10^{-12}, and every joint fit converged.

Results of study E. All five registered hypotheses of Table 30 hold.

Sharing the tilt across offers along the framework's dose curve improves the trial by 0.152 [0.146, 0.158] at one week.

At the same parameterization, the framework's regret is below that of the joint fit by 0.055 [0.051, 0.059], 0.052 [0.048, 0.056] and 0.060 [0.053, 0.066] at one, two and four weeks, and its cost tables are more accurate, with relative cost-table error 0.076 against 0.090 at one week, a difference of 0.014 [0.014, 0.015].

At one week, sharing across offers accounts for 73.5% [72.1, 74.8] of the framework's advantage over the tilted trial and placement for 26.5% [25.2, 27.9]. The placement share in Table 17 grows with the pilot length as the per-offer fits improve.

Table 17: Study E: relative regret within 20% of the budget, 288 response laws × 8 replicates × 60 requests × 3 budgets; share of the difference to the tilted trial due to placing the observations at three offers, with 95% law-bootstrap intervals.

Method	1 wk	2 wk	4 wk
Framework, pilot at three offers	**0.198**	**0.154**	**0.123**
Joint fit of the trial, framework's dose curve	0.253	0.206	0.183
History-anchored tilt of the trial, per offer	0.405	0.289	0.211
Same-calendar RCT	0.479	0.337	0.241
Placement share	0.265 [0.252, 0.279]	0.382 [0.361, 0.405]	0.683 [0.623, 0.745]

By the risk components of Table 18 at one week, the framework's plans fall below the band for 3.4% of requests against 6.5% for the joint fit and above it for 4.1% against 4.4%, they leave it by less, 3.5% against 4.6% of the budget, and its valid plans lose less value, 13.6% against 16.5%.

Table 18: Risk components at one week within 20% of the budget, study E, 288 response laws, law means. Every arm issues a plan for more than 99.99% of the requests. Below/above: share of requests whose plan has true expected spend below/above the band. Excursion: mean distance outside the band, in units of the budget, over the plans that leave it. Valid-plan loss: value lost by the plans inside the band relative to the band optimum.

Method	Below	Above	Excursion	Valid-plan loss
Framework	**0.034**	**0.041**	**0.035**	**0.136**
Joint fit of the trial	0.065	0.044	0.046	0.165
Dose interpolation, same information	0.044	0.045	0.041	0.149
History-anchored tilt of the trial, per offer	0.179	0.010	0.070	0.266
Same-calendar RCT	0.272	0.010	0.082	0.288

Weight diagnostics. For every group, rule and non-zero offer, where at the unpiloted ones among the eight offers the tilt follows the dose curve, the study records the effective sample size $\mathrm{ESS} = (\sum_i w_i)^2 / \sum_i w_i^2$ of the tilt over the $N = 4{,}096$ stored weeks. It also records $N \max_i w_i$, whether the moment target lies in the coordinate-wise range, or box, of the stored weeks' features, a necessary condition for lying in their convex hull, whether the fit reaches its tolerance, and the condition number of the unweighted covariance of the stored weeks' features. Percentiles below are over all such cells at one week, independent condition.

- ESS/N: median 0.951, 5th percentile 0.600, 1st percentile 0.422.
- $N \max_i w_i$: median 2.05, 95th percentile 12.6, 99th percentile 26.4, so even at the 99th percentile no stored week carries more than 0.65% of the mass.
- The target lies outside the feature box in 0% of the cells, and no fit fails to reach its tolerance.
- The covariance condition number has median 7.8, 95th percentile 31.3 and 99th percentile 47.4.

Both registered diagnostic hypotheses hold. Within a law, the ESS deficit $1 - \mathrm{ESS}/N$ is rank-correlated with the cost-table error of the framework, with a Spearman correlation of 0.415 [0.400, 0.431] when tied values get their average rank. Fitting the tilt to exact pilot moments on 1,024 rather than 16,384 stored weeks raises the relative cost-table error by 0.006 [0.006, 0.006]. These are heuristic stability diagnostics: fits converge, weights are not concentrated, and the ESS deficit tracks the cost-table error. They do not check the interior-hull and eigenvalue conditions of Lemma 6 along the parameter path, which remain sufficient conditions of the analysis; a condition number is not an eigenvalue bound.

Study D: requests that change behaviour. It draws 360 laws: 216 window-dependent laws in three families, 72 laws whose response does not depend on the window, and 72 null laws. In the window-dependent families, a short paid window concentrates effort on its days, riders shift some

activity away from unpaid days, and less active riders respond more. A request may also target the riders above or below the group's median of the previous no-offer week, whose response differs from the group's.

- *Direct reuse* keeps the full-week, group-level calibration of Appendix D for every request.
- *Request-aware reuse* runs the pilot with a random paid window per rider. It then fits a window-relative tilt on four features, the completions and active days inside and outside the paid window, pooled over the pilot riders' windows. Stored weeks and pilot riders are conditioned on the request's selection. The tilt is fitted once per origin for every group, rule, selection and piloted offer, and reused by every request.

The objective is the uplift in weekly completions, and the payment follows the request's window and rule. The truth is exact.

Table 19: Study D: relative regret within 20% of the budget on the 216 window-dependent laws, 8 replicates $\times$ 60 requests $\times$ 3 budgets. Selected: requests that target a subpopulation chosen on the previous week.

Method	1 wk	2 wk	4 wk	Selected, 1 wk
Request-aware reuse	**0.380**	**0.329**	**0.297**	**0.395**
Direct reuse of the full-week calibration	0.576	0.584	0.596	0.742

Results of study D. All five registered hypotheses of Table 30 hold.

When the window changes behaviour, request-aware reuse lowers regret against direct reuse by 0.195 [0.189, 0.202], 0.255 [0.249, 0.261] and 0.300 [0.293, 0.306] at one, two and four weeks. For requests that target a selected subpopulation the reduction is 0.347 [0.338, 0.356] at one week.

On all 288 response laws, request-aware reuse has lower regret than the same-calendar trial by 0.129 [0.123, 0.136] at one week.

The reused tilt returns the same answers as a tilt re-fitted for every request in all 18,432 comparisons at each pilot length. With preparation charged it is an estimated 3.13$\times$ [3.09, 3.16] faster over 60 requests at one week, and 3.07$\times$ and 3.09$\times$ faster at two and four weeks; preparation and all 60 reused requests are timed, and the time without reuse is 60 times the mean of the eight re-fitted requests timed per replicate.

The study makes the reuse rule explicit. A cached calibration is valid for the windows and populations on which the response law was measured. When the paid window or the selection changes behaviour, the cache is keyed by window-relative features and by the selection, and the pilot must randomize windows. Its cost is the pilot's calendar weeks, which are charged above.

Study F: value retained by population reduction. Study F prices the reduction stage against a common reference optimum. It draws 3,072 laws from the rider-level generator, 1,024 in each response family, and scores every method with the exact per-type tables, so estimation error is absent. Each request takes a window, one rule and two of the six groups, with three budgets; every budget binds. A rider's history statistic is its total completions over the previous four no-offer weeks. The reference is a *type-informed continuous-population optimum*: it knows every rider's latent type and treats each group as a continuous population, so it may split the population mass of one type across several offers. It is a linear program over types, computed exactly as the sup-convolution of the types' upper concave hulls, and every layered plan is a feasible point of it. It therefore holds more information than any layered plan, since it sees latent types instead of the four-week history, and more allocation authority, since it divides type masses instead of assigning one offer per layer, and it is not the integer optimum over a finite set of riders. The framework's reduction cuts each group into L equal-share quantile layers of the history statistic with one offer per layer, and the layered optimum is found by exact enumeration. The endpoint is the retained share: the value of a method summed over a law's requests, divided by the reference optimum summed over the same requests.

Table 20: Study F: share of the type-informed continuous-population optimum retained by L history-quantile layers per group and by five equal-width bins of the same history statistic, within 20% of the budget; 3,072 laws, 60 requests $\times$ 3 budgets each; 95% law-bootstrap intervals. By family: saturating response on both channels, intensity only, threshold response near the goal.

Layers per group	All laws	Both channels	Intensity	Threshold
1, group-level plan	0.825 [0.820, 0.831]	0.938	0.944	0.594
2	0.900 [0.897, 0.904]	0.966	0.971	0.764
3	0.926 [0.923, 0.928]	0.974	0.978	0.825
4	0.938 [0.936, 0.940]	0.977	0.982	0.856
5	**0.946 [0.944, 0.948]**	**0.979**	**0.984**	**0.875**
5 equal-width bins	0.947 [0.946, 0.949]	0.980	0.984	0.878

Results of study F. Five history layers retain 94.6% [94.4, 94.8] of the type-informed continuous-population optimum, against 82.5% [82.0, 83.1] for one group-level plan: 0.120 [0.116, 0.125] more, which recovers 68.9% [68.8, 69.0] of the group-level loss, as Table 20 shows.

Every added layer raises the retained share: five layers retain 0.046 [0.044, 0.047] more than two, and the fourth and fifth layers add 0.013 [0.012, 0.013] and 0.008 [0.007, 0.008].

The number of layers carries the gain: five equal-width bins of the same statistic retain 94.7% [94.6, 94.9], and 98.0%, 98.4% and 87.8% by family.

Retention is highest when the response is smooth, 97.9% and 98.4% with five layers, and lowest under a threshold response near the goal, 87.5%, where the plan that pays depends most on a rider's distance to the tier.

No layered plan abstains in this band and every budget binds, so the gap measured here is that of giving one offer to each history layer instead of dividing each latent type's mass with knowledge of the types; estimation error, which studies A–E price, comes on top of it.

Study G: departures from the tilt family. Study G maps the range of the advantage of study E when the response law departs from the tilt family. It reuses the 288 response laws of study C with fresh requests and data, eight replicates, 60 requests and a one-week pilot, with the observation counts of 486 randomized rider-weeks per group for the pilot and for the trial, the 4,096 stored weeks and the exact enumeration of study E. Level M0 is study E's generator; three sweeps each change one condition, at three strengths fixed before the runs.

- *Response tail*, strengths T1–T3. Only a share $\pi = 1/2, 1/4, 1/8$ of the rider-weeks responds to an offer, with both response amplitudes multiplied by $1/\pi$; the other rider-weeks behave as without an offer.
- *History-support gap*, strengths S1–S3. The top share $\delta = 0.10, 0.25, 0.50$ of the no-offer weeks, ranked by weekly completions, is removed from the stored history, leaving 4,096 weeks, so part of the trajectories that an offer induces lies outside the history that the tilt reweights. Riders, pilot, trial and truth are unchanged.
- *Curvature at untested doses*, strengths K1–K3. The dose function f of every response family becomes $f + \kappa h$ with $\kappa = 0.25, 0.5, 1$, where h is the quadratic hump that is zero at no offer and at the three piloted offers and one midway between them.

Every arm is that of study E. A *guarded framework* uses the same information as the framework: for every request and group it compares, at each piloted offer, the pilot's mean payment on the request's window with the framework's predicted mean payment, standardized by the predicted standard error; if the largest of the three values exceeds 3, a threshold fixed before the runs, that group's table is replaced by dose interpolation. When a pilot or trial target lies outside what the stored weeks can reach, as under support gaps, the fits stop at iteration budgets fixed before the runs, 100 Newton steps for the framework's tilt and 500 iterations for the joint fit, and the arms use that iterate; Table 22 reports how often this happens. The registered family has 18 one-sided tests, the framework minus the joint fit and the framework minus the same-calendar trial at each of the nine strengths, with Holm

step-down; the advantage range of a sweep is the set of strengths at which its test holds. The crossover of a sweep is the strength at which the law-mean contrast reaches zero, by linear interpolation between tested strengths on $\log_2(1/\pi)$, δ and κ, with a percentile interval over the same bootstrap draws.

A second registered study on the same laws, data and requests separates three sources of an unconverged fit, reported in Table 23. *Reachability*: a tilt with positive weights meets a moment target exactly when the target lies in the interior of the convex hull of the stored weeks' features, the normalized weekly completions and active days; this is decided exactly in integer arithmetic on the feature lattice and checked by a linear program, with no disagreement in any condition. The joint fit has a finite maximizer exactly when the size-weighted trial means lie in the interior of the corresponding weighted Minkowski sum of hulls, decided by a linear program. *Iteration budgets*: the framework's fit is continued to 500 and 2,000 Newton steps and the joint fit to 2,000 and 5,000 iterations, under the output rule above. *A uniform rule for unreachable targets*: every target outside the hull is projected onto the hull shrunk by 2% toward the history mean, reachable targets are kept, and every fit, of the framework, the per-offer tilt and the joint fit, is run by a damped Newton method to a tolerance of 10^{-9}. Its registered family has 16 one-sided tests with Holm step-down, the framework minus the joint fit under the uniform rule and with the larger budgets of 2,000 and 5,000 steps, at M0 and at the seven strengths where the corresponding test of study G holds. The same runs add guard variants, reported descriptively: thresholds 2 and 5 for the pilot-payment check, the check applied to the projected fit, and a *reachability guard* with the same information as the framework, which replaces a group's table for a rule by dose interpolation whenever one of its three pilot targets lies outside the hull by the exact test.

Table 21: Study G: relative regret within 20% of the budget, one-week pilot, 288 response laws $\times$ 8 replicates $\times$ 60 requests $\times$ 3 budgets, at every strength inside the advantage range of both comparisons; differences with 95% law-bootstrap intervals. M0: generator of study E, fresh data.

	Regret			Framework minus	
Condition	**Framework**	Joint fit	Same-calendar	joint fit	same-calendar trial
M0	**0.197**	0.254	0.481	−0.056 [−0.060, −0.052]	−0.283 [−0.291, −0.276]
Tail, $\pi = 1/2$	**0.205**	0.260	0.491	−0.055 [−0.058, −0.052]	−0.286 [−0.293, −0.279]
Tail, $\pi = 1/4$	**0.222**	0.275	0.514	−0.054 [−0.057, −0.050]	−0.293 [−0.299, −0.286]
Tail, $\pi = 1/8$	**0.249**	0.314	0.551	−0.065 [−0.069, −0.061]	−0.302 [−0.307, −0.297]
Support gap, $\delta = 0.10$	**0.236**	0.282	0.481	−0.046 [−0.049, −0.043]	−0.244 [−0.250, −0.238]
Support gap, $\delta = 0.25$	**0.333**	0.349	0.481	−0.016 [−0.021, −0.011]	−0.148 [−0.157, −0.138]
Curvature, $\kappa = 0.25$	**0.297**	0.321	0.435	−0.024 [−0.030, −0.018]	−0.138 [−0.148, −0.127]

Table 22: Study G at one week, law means: fit diagnostics, the pilot-payment check and the regret, within 20% of the budget, of the framework, the guarded framework, the reachability guard and dose interpolation used throughout. Moment residual: share of replicate–group–rule units whose largest moment residual at the piloted offers exceeds 10^{-6} after the tilt's 100 Newton steps; joint fit converged: share of joint fits that converge within 500 iterations; flagged: share of request–group cells in which the pilot-payment check exceeds 3; the reachability guard falls back in a request–group cell when a pilot target of the group lies outside the hull of the stored weeks; ESS/N: effective sample size of the tilt over the 4,096 stored weeks. Under the curvature sweep the data at the piloted offers are those of M0, so the tilt diagnostics and the flag share equal M0's.

	M0	$\pi = 1/2$	$\pi = 1/4$	$\pi = 1/8$	$\delta = 0.10$	$\delta = 0.25$	$\delta = 0.50$
Moment residual $> 10^{-6}$, %	0	0	0	0	0.096	39.8	99.97
Joint fit converged, %	100	100	100	100	99.96	69.0	0.063
Flagged, %	0.015	0.011	0.036	0.068	1.0	19.7	85.0
Reachability guard falls back, %	0	0	0	0	0.084	39.4	99.96
ESS/N	0.864	0.866	0.879	0.901	0.596	0.219	0.030
Regret: framework	0.197	0.205	0.222	0.249	0.236	0.333	0.809
guarded framework	0.197	0.205	0.222	0.249	0.235	0.286	0.254
reachability guard	0.197	0.205	0.222	0.249	0.236	0.261	0.245
dose interpolation	0.219	0.225	0.237	0.255	0.240	0.245	0.245

Table 23: Study G: reachability of the moment targets and convergence of the fits at one week, in %. Units: replicate–group–rule; a pilot target is reachable when it lies in the interior of the convex hull of the stored weeks' features by the exact test; a joint problem is solvable when its likelihood has a finite maximizer. Converged: largest moment residual at most 10^{-9} for the tilt, or score per observation at most 10^{-9} for the joint fit. Every tail and curvature strength has the reachability and convergence of M0.

	M0, tails, curvature	$\delta = 0.10$	$\delta = 0.25$	$\delta = 0.50$
Pilot targets outside the hull	0	0.031	19.9	98.6
Units with all three pilot targets reachable	100	99.91	60.6	0.036
Joint problems without a finite maximizer	0	0.036	31.0	99.94
Tilt: reachable, not converged after 100 steps	0	0.010	0.43	0.005
after 500 or 2,000 steps	0	0	0	0
moment residual $> 10^{-6}$ after 2,000 steps	0	0.087	39.1	99.96
Joint fit: solvable, not converged after 500 iterations	0	0	0.0024	0
after 2,000 or 5,000 iterations	0	0	0	0
converged after 5,000 iterations	100	99.96	69.1	0.063
Uniform rule: all fits converged	100	100	100	100

Results of study G. *Advantage range*, in Table 21 with tests in Table 30. The framework's regret is below that of the joint fit and of the same-calendar trial at every tested response tail, down to a responder share of 1/8, by 0.054 to 0.065 and by 0.286 to 0.302. At a support gap of 10%, where 0.031% of the pilot targets and 0.036% of the joint problems lie beyond the reach of the stored weeks in Table 23, it is below the joint fit by 0.046 and below the same-calendar trial by 0.244. At gaps of 25% and 50% the moment targets of many groups lie beyond what the stored weeks reach, with moment residual above 10^{-6} in 39.8% and 99.97% of units and joint fits converged in 69.0% and 0.063%, so these contrasts compare the outputs after 100 and 500 steps under the registered rule: the framework's output is below the joint fit's by 0.016 and 0.078, and below the same-calendar trial by 0.148 at the 25% gap. At curvature $\kappa = 0.25$ it is below both, by 0.024 [0.018, 0.030] and 0.138 [0.127, 0.148].

Reachability, not the iteration budget, explains the unconverged fits, as Table 23 shows. No pilot target is out of reach at M0 or under any tail or curvature strength. At gaps of 10%, 25% and 50% the exact test places 0.031%, 19.9% and 98.6% of the pilot targets outside the hull, and 0.036%, 31.0% and 99.94% of the joint problems have no finite maximizer. Every unit whose three pilot targets are reachable meets them within 500 Newton steps, with largest moment residual 2.1×10^{-14}, and every joint problem with a finite maximizer converges within 2,000 iterations. Every fit that remains unconverged has an unreachable target or no finite maximizer, and beyond 500 steps for the framework and 2,000 iterations for the joint fit further steps leave the converged shares unchanged.

Converged fits under one rule, with tests in Table 31. Under the uniform rule every fit converges to 10^{-9} in every condition, the framework's within at most 346 Newton steps. The framework's regret is then below the joint fit's at M0, under all three tails, at gaps of 10% and 25% and at $\kappa = 0.25$, by 0.056, 0.055, 0.054, 0.065, 0.046, 0.015 [0.010, 0.019] and 0.024 [0.018, 0.030]. With the larger budgets under the output rule of study G it is below the joint fit in all eight registered conditions, by 0.007 [0.002, 0.013] at the 25% gap; at the 50% gap, 0.098 [0.093, 0.102], both fits pursue targets that no reweighting of the stored weeks reaches, so this value compares outputs.

Crossover points, reported descriptively. Against the same-calendar trial the advantage extends to a support gap of $\delta = 0.327$ [0.323, 0.331]. Along the curvature sweep it extends to $\kappa = 0.371$ [0.337, 0.408] against the joint fit and to $\kappa = 0.430$ [0.415, 0.445] against the same-calendar trial. Along the tail sweep it covers the whole tested range in every bootstrap draw, and so does the output comparison with the joint fit along the support sweep.

Reproduction. On fresh data, M0 reproduces study E: the framework minus the joint fit is -0.056 [-0.060, -0.052], against -0.055 in study E.

A check fixed in advance detects support gaps, as Table 22 shows. The pilot-payment check flags 0.015% of the cells at M0 and at most 0.07% under the tail sweep, and 1.0%, 19.7% and 85.0%

at support gaps of 10%, 25% and 50%. The guarded framework lowers regret by 0.047 [0.043, 0.051] at $\delta = 0.25$ and by 0.555 [0.538, 0.572] at $\delta = 0.50$, where its regret of 0.254 is below the same-calendar trial's 0.480, a difference of -0.226 [-0.233, -0.220]; without a support gap it leaves regret unchanged to within 0.001. The guard is a trade-off between keeping the framework's accuracy under normal conditions and the cost of its fallback: with full support it keeps the framework's regret, 0.021 [0.020, 0.023] below using dose interpolation throughout; at gaps of 25% and 50%, where only the flagged groups fall back, its regret differs from using dose interpolation throughout, which gives 0.245 at both gaps, by 0.041 [0.037, 0.045] and 0.009 [0.008, 0.010]. Over the whole sweep the guard keeps 99.9% [99.8, 100.0] of the framework's advantage over dose interpolation at M0 and stays below dose interpolation at every tail and curvature strength and at the 10% gap, by 0.002 to 0.020. The threshold sets the balance: at 2 and at 5 the guard is below dose interpolation by 0.021 at M0 and differs from it by 0.028 and 0.056 at the 25% gap and by 0.003 and 0.038 at the 50% gap. The reachability guard uses the same information as the framework and falls back only on groups with an unreachable pilot target: it leaves the framework's regret unchanged at M0 and under every tail and curvature strength, where it never falls back, differs from dose interpolation by 0.017 [0.014, 0.019] at the 25% gap, and equals it to within 10^{-4} at the 50% gap, where 99.96% of cells fall back. Every guard is below the same-calendar trial at M0, under every tail and at every support gap, the pilot-payment check by 0.194 and 0.226 at the 25% and 50% gaps. The weight diagnostic moves with the support gap in the same way: ESS/N falls from 0.86 to 0.60, 0.22 and 0.03, and stays between 0.86 and 0.90 under the tail sweep.

Pilot design for curvature. A bend of the dose response between piloted offers leaves the pilot's data at the piloted offers unchanged, so it is covered by the pilot design rather than by a diagnostic: the piloted offers should include, or lie next to, the doses that plans use.

Study H: finite riders and a same-information reference. The reference of study F divides type masses across offers and knows the latent types. Study H separates these two advantages on 768 of the laws of study F, 256 per response family, with the same requests, budgets and bands. For $N = 5$, 10 and 20 riders per group, with $N = 20$ primary, each rider's latent type and four-week history statistic are drawn once per law and group, and every rider receives exactly one offer. Type scoring uses each rider's type tables; history scoring uses the posterior mixture of the type tables given the rider's history, which is what a planner who sees the history but not the type expects. The references are the type-informed continuous-population optimum of study F, the continuous-population optimum over units of one group and history value with posterior tables, which has the same information as the layers, and the exact integer optimum over the finite riders under either scoring, together with its linear programming (LP) relaxation, in which riders with equal tables may be split across offers. The integer optimum is computed by an exact Pareto recursion over riders under the cap, then under the floor, and otherwise by complete enumeration. At $N = 5$ every integer optimum agrees with complete enumeration of all 9^{10} assignments, with a largest difference of 7.1×10^{-15}. Every integer problem of the primary band is solved exactly; of the 1,658,880 problems over both bands, one in the stricter 10% band, for one law with ten riders under history scoring, is recorded as unsolved. Layers are equal-count history quantiles within the group, one offer per layer, solved exactly. The registered family has four one-sided tests with Holm step-down.

Table 24: Study H: share of four reference optima retained by L history-quantile layers per group within 20% of the budget; 768 laws, 60 requests $\times$ 3 budgets each. Continuum: divisible population masses; integer: 20 riders per group, one offer per rider. Type-informed: scored with the latent types; same information: scored with the posterior given the history.

	Continuum		Integer, 20 riders per group	
Layers per group	Type-informed	Same information	Type-informed	Same information
1, group-level plan	0.825	0.851	0.834	0.858
2	0.900	0.932	0.910	0.937
3	0.925	0.959	0.938	0.965
4	0.938	0.973	0.951	0.977
5	**0.946**	**0.981**	**0.960**	**0.984**
95% interval	[0.942, 0.949]	[0.979, 0.982]	[0.957, 0.963]	[0.982, 0.985]

Results of study H. All four registered hypotheses of Table 30 hold.

Finite riders. With one offer per rider, the exact integer optimum lies on average 1.3%, 0.49% and 0.19% below the LP relaxation of the same finite problem at 5, 10 and 20 riders per group under type scoring, and 1.2%, 0.47% and 0.18% under history scoring. Each law's gap pools its requests, and these values are means over laws; at 20 riders the largest law's gap is 0.68%. The LP relaxation is the reference for indivisibility only; it differs from the type-informed continuum optimum of study F. At 20 riders five layers retain 96.0% [95.7, 96.3] of the type-informed integer optimum, 0.126 [0.118, 0.135] more than the 83.4% of one plan per group.

Same information. Five layers retain 98.4% [98.2, 98.5] of the same-information integer optimum at 20 riders, 0.034 [0.032, 0.035] above 0.95, and 98.1% [97.9, 98.2] of the same-information continuum optimum. Measured against the same information, the retained share is 0.024 [0.022, 0.025] higher than against the type-informed integer optimum, and 0.035 [0.033, 0.037] higher in the continuum.

Decomposition of study F's loss. On these laws five layers lose 0.054 [0.051, 0.058] of the type-informed continuum optimum. Knowing the latent types accounts for 0.036 [0.034, 0.039] of it, 66.6% [66.0, 67.2], and layering at the same information for 0.018 [0.017, 0.019].

By family. At 20 riders five layers retain 99.5% and 99.6% of the same-information integer optimum under the two smooth response families and 95.9% under a threshold response near the goal.

Study I: number of layers at a fixed pilot and history. Studies F and H score layered plans with exact tables, and studies A–E estimate the response at the group level. Study I runs the whole finite-data pipeline with $L = 1$, 3 and 5 history layers per group at the same totals, so that more layers bring their approximation gain but split the pilot and the stored history into smaller cells. It uses 1,536 laws of study F, 512 per response family, two replicates per law, and study F's 60 requests with three budgets each. The resources per group are $n = 135$, 270, 540, 1,080, 2,160, 4,320 or 8,640 randomized rider-weeks, where $n = 540$ is about the one-week pilot of studies B–E, and $H = 8n$ stored rider records, each a rider's four-week history statistic and one further no-offer week. The pipeline cuts the stored records into L equal-count layers of the history statistic, spreads the n rider-weeks as evenly as possible over layers, rules and the three piloted offers, tilts each layer's stored weeks to that layer's pilot moments with 100 Newton steps, sums each request's rule over the stored weeks, allocates exactly over the 9^{2L} portfolios within 20% of the budget, and is scored on the exact tables of the finite layers. The endpoint is the retained share of study F's type-informed continuous-population optimum, and the net gain of L layers is $D(L, n) = \text{retained}(L, n) - \text{retained}(1, n)$. Per law it splits exactly as

$$D(L, n) = G(L) - E(L, n),$$

with the approximation gain $G(L)$ of L population layers on the exact tables of study F and the estimation error $E(L, n)$ that L layers add; E splits further into a part due to placing riders in layers from finite records and a part due to estimating the response. The registered family has 14 one-sided tests, $D(L, n) > 0$ for $L = 3$ and 5 at each level, with Holm step-down; the resource range of L is the set of levels at which its test holds.

Table 25: Study I: net gain of three and five history layers over one plan per group with the same pilot and history, the added estimation error and five minus three layers, within 20% of the budget; 1,536 laws × 2 replicates × 60 requests × 3 budgets, 95% law-bootstrap intervals. Approximation gains on exact tables: $G(3) = 0.100$ [0.096, 0.105], $G(5) = 0.121$ [0.115, 0.126].

n	$D(3, n)$	$D(5, n)$	$E(3, n)$	$E(5, n)$	Five minus three
135	0.060 [0.057, 0.063]	0.026 [0.023, 0.030]	0.040	0.094	−0.034
270	0.081 [0.078, 0.084]	0.077 [0.074, 0.080]	0.019	0.043	−0.004
540	0.086 [0.083, 0.089]	0.095 [0.091, 0.098]	0.014	0.026	0.009 [0.007, 0.011]
1,080	0.086 [0.083, 0.090]	0.102 [0.098, 0.106]	0.014	0.018	0.016 [0.014, 0.018]
2,160	0.079 [0.076, 0.083]	0.099 [0.095, 0.103]	0.021	0.022	0.020 [0.018, 0.021]
4,320	0.073 [0.069, 0.077]	0.092 [0.087, 0.096]	0.027	0.029	0.018 [0.016, 0.020]
8,640	0.066 [0.062, 0.070]	0.079 [0.075, 0.084]	0.034	0.041	0.013 [0.011, 0.016]

Results of study I. All 14 registered hypotheses of Table 31 hold.

Resource range. With the same pilot and history, three and five layers retain more of the reference optimum than one plan per group at every level from 135 to 8,640 randomized rider-weeks per group, by 0.060 to 0.086 and by 0.026 to 0.102, as Table 25 shows; the resource range of both is all seven levels.

Approximation gain against added estimation error. The approximation gain, 0.100 for three layers and 0.121 for five, exceeds the added estimation error at every level; the added error is smallest at $n = 1{,}080$, 0.014 and 0.018, and largest at $n = 135$, 0.040 and 0.094. Almost all of it comes from estimating the response in smaller cells: the part due to placing riders in layers from finite records is below 10^{-4} in absolute value at every level, and the decomposition holds per law to 10^{-16}.

Three or five layers. Three layers are the best choice at $n \leq 270$ and five layers from $n = 540$ on; five minus three layers crosses zero at $n = 329$ [298, 357] rider-weeks per group, below the one-week pilot.

Against the joint fit. With the full pipeline the framework retains more than the joint fit of a nine-offer trial with the same rider-weeks at every number of layers and level, by 0.032 to 0.068 with one layer, 0.047 to 0.122 with three and 0.048 to 0.135 with five.

Scope of the response evidence. The response evidence of Section 5 and of studies A–G and I comes from controlled response families: saturating, intensity-only and threshold responses in the 32-type generator and in the rider-level generator, with rule-dependent responses, shared calendar shocks and behaviour-changing windows and subpopulations. Within these families the correction lowers regret at every pilot length tested. Its use elsewhere rests on the conditions priced by the synthetic term of Proposition 1: the no-offer history must cover the trajectories that the offer induces, the two matched features must carry the response, and the dose curve is interpolated between piloted offers. Study G maps these conditions: the advantage holds over the tested response tails and for curvature between piloted offers up to $\kappa \approx 0.37$; under support gaps it holds against the same-calendar trial up to a quarter of the support removed and, with converged fits and one rule for unreachable targets, against the joint fit up to a quarter. An exact reachability test finds the groups whose targets the stored weeks cannot reach, and a guard that falls back to dose interpolation on them keeps the framework's regret elsewhere. Study I shows that the value of more layers survives estimation from finite pilots and histories at every tested resource level. Rider records test pricing and computation on logged completions, not behavioural response.

F Learned pipelines with the same information and one frozen operating policy

Study J: learned response pipelines with the same information and cache. Study J asks whether the framework's history-anchored tilt beats learning pipelines that a practitioner would build from the same data, namely the stored no-offer weeks and the pilot at three offers, when every pipeline prepares once per origin, caches every table it computes and feeds the same exact integration and allocation. It uses study E's generator in the independent condition of study C, the 288 response laws of study C, 8 replicates, 60 requests with three budgets each, pilots of one, two and four weeks at offers 2, 5 and 8 and 4,096 stored weeks per group, all drawn from fresh streams. Every arm answers requests through a table cache keyed by group, rule and window, and allocation enumerates all portfolios.

- *Learned density ratio* by probabilistic classification. Per group and rule, a multinomial logistic classifier takes the stored weeks as reference class and one class per piloted offer, on 19 trajectory features: per-day completions and activity indicators, weekly completions and active days, their squares and their product, all standardized on the stored weeks. The L2 penalty on the slopes is chosen from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$ by 3-fold stratified cross-validated log-loss, and the fit is by Newton's method. The class odds give the density ratio of each piloted offer to no offer, which reweights the stored weeks; the slope vectors are interpolated piecewise linearly in the reward increment through zero at no offer, as the framework's λ is. The same learner on the framework's two features is reported descriptively.
- *Response-model pipeline*. For each group, rule and window, the expected payment and uplift at the observed offers, with the zero-increment tariff settled on the stored weeks and piloted

offers on their cells, are made non-decreasing in the reward increment by weighted isotonic regression with the baseline fixed, and are interpolated at the other offers by a monotone piecewise-cubic curve; the resulting cost and uplift tables feed the same allocation. The baseline has zero uplift but may have positive payment: its cost is the historical mean under the same tariff function used by every arm; zero reward increment does not mean zero settlement.

Dose interpolation, the same-calendar trial and the joint fit of the trial are those of studies E and G. Time is measured per replicate, single-threaded, as preparation, that is the framework's tilt fits or the classifier's cross-validation and fits, plus answering the 60 requests with the cache. The registered family has six one-sided tests, the framework minus each learned pipeline at each pilot length, with Holm step-down at level 0.025.

Response-model anchor. Amendment E14-2 of the registration, which follows the registered run and analysis of study J, sets the payment of the response-model pipeline at zero reward increment to the shared tariff; in the registered run it was settled as 0. The pipeline is replayed deterministically with this anchor on the registered laws, replicates, requests and random streams, and the family of six tests is evaluated on this replay; the results reported here are those of the replay, and its timings are not reported. All other arms keep the results of the registered run.

Table 26: Study J: relative regret of learned pipelines with the framework's information and cache, 288 response laws $\times$ 8 replicates $\times$ 60 requests $\times$ 3 budgets, with planning, optimum and scoring within 20% or 10% of the budget. Time is preparation plus 60 requests, single-threaded, as the median over laws of the median over replicates with a one-week pilot, and as the geometric mean over laws of its ratio to the framework's; a dash denotes a runtime that is not reported.

	Within 20% of budget			Within 10%			Time, 1 wk	
Method	1 wk	2 wk	4 wk	1 wk	2 wk	4 wk	s	ratio
Framework	**0.196**	**0.153**	**0.122**	**0.284**	**0.199**	**0.148**	1.10	1.00
Learned density ratio, 19 features	0.552	0.537	0.466	0.629	0.579	0.489	42.71	35.96
the same learner, two features	0.333	0.264	0.203	0.384	0.293	0.218	8.88	8.00
Response-model pipeline	0.216	0.169	0.133	0.312	0.215	0.158	–	–
Dose interpolation	0.218	0.171	0.135	0.319	0.224	0.164	0.81	0.77
Same-calendar trial	0.481	0.338	0.242	0.767	0.616	0.421	–	–
Joint fit of the trial	0.252	0.208	0.186	0.386	0.281	0.225	–	–

Table 32 reports all six registered tests, and all six hold under the joint Holm correction. At one week the learned density ratio's excess regret over the framework is 0.356 [0.347, 0.364], and that of the response-model pipeline is 0.020 [0.018, 0.022]. Table 26 gives both budget bands at all three pilot durations. The framework's one-week regret is thus 9.3% below the response model's, 0.196 against 0.216. Restricting the classifier to the framework's two features lowers its one-week regret by 0.219 [0.212, 0.227], and it remains 0.136 [0.129, 0.143] above the framework. The classifier's cross-validation makes its preparation plus 60 requests 36.0$\times$ [35.1, 36.8] the framework's time.

Study K: one frozen operating policy. Study G examined the framework and its guards one departure at a time. Study K fixes one operating policy, OP, that combines the base tilt, a validation check at untested offers, a supplementary refit, a reachability check and a payment check. It pays for the validation out of the same randomized rider-weeks as the framework's pilot and runs unchanged over all ten conditions of study G, with the generator of study G, the 288 response laws, fresh data, 8 replicates, 60 requests with three budgets and a one-week pilot. Of the $9\lfloor n/27 \rfloor$ pilot weeks per group and rule, $2\lfloor n/27 \rfloor$ go to each piloted offer 2, 5 and 8 and $\lfloor n/27 \rfloor$ to each validation offer 1, 3 and 6, the midpoints of the intervals between no offer and the piloted offers. Per group and rule:

1. *Reachability*: if a moment target that a fit uses lies outside the interior of the convex hull of the stored weeks' features by the exact test of study G, the group and rule use the fallback for every request.
2. *Base*: the tilt at the three piloted offers, with λ piecewise linear in the reward increment through zero at no offer.
3. *Validation*: at each validation offer v, $T_v = n_v(m_v - \mu_v)^\top S_v^{-1}(m_v - \mu_v)$, where m_v holds the cell's mean normalized weekly completions and active days and μ_v, S_v are their mean

and covariance under the base law at v; the unit is flagged when $\max_v T_v$ exceeds the χ^2_2 quantile at $1 - 0.05/3$, which is 8.19.

4. *Supplementary refit* for flagged units: the tilt at all six observed offers, with λ piecewise linear through zero and the six fitted values; step 1 applies to the six targets.

5. *Payment check* per request: at every offer the fitted law uses, the z-value of the cell's mean payment on the request's window against the law's mean payment, standardized by the law's payment variance over the cell size; if the largest exceeds 3, the group's table for that request is the fallback's.

The fallback is dose interpolation of the stored weeks and OP's six cells with mixture weights piecewise linear in the reward increment; it is OP's same-information alternative. The registered family has 23 one-sided tests with Holm step-down at level 0.025: OP minus the framework at M0 with non-inferiority margin 0.02; OP minus the same-calendar trial in each of the ten conditions; OP minus its fallback with margin 0.02 in each condition; and OP minus the framework under curvature 0.5 and 1. A non-inferiority p-value is the bootstrap share of draws with mean difference at or above the margin.

Table 27: Study K: one frozen operating policy OP run unchanged over the ten conditions of study G, one-week pilot, 288 response laws × 8 replicates × 60 requests × 3 budgets; relative regret within 20% of the budget. Fallback: OP's same-information fallback, dose interpolation of the stored weeks and OP's six cells. Flagged: share in % of replicate, group and rule units in which the validation check at the untested offers fires; unreachable: share in % of units in which a moment target lies outside the hull of the stored weeks. Differences with 95% law-bootstrap intervals.

Condition	**OP**	Framework	Fallback	Same-cal.	Joint fit	OP minus framework	Flagged	Unreach.
M0	**0.250**	0.198	0.383	0.478	0.253	+0.052 [+0.048, +0.055]	12.3	0.0
Tail, $\pi = 1/2$	**0.263**	0.205	0.392	0.490	0.257	+0.057 [+0.054, +0.060]	14.2	0.0
Tail, $\pi = 1/4$	**0.293**	0.221	0.411	0.513	0.277	+0.072 [+0.068, +0.075]	16.5	0.0
Tail, $\pi = 1/8$	**0.333**	0.247	0.441	0.547	0.311	+0.086 [+0.082, +0.090]	18.5	0.0
Support gap, $\delta = 0.10$	**0.330**	0.238	0.382	0.478	0.280	+0.093 [+0.087, +0.098]	47.0	0.5
Support gap, $\delta = 0.25$	**0.371**	0.334	0.382	0.478	0.349	+0.037 [+0.028, +0.046]	49.1	55.6
Support gap, $\delta = 0.50$	**0.381**	0.808	0.381	0.478	0.888	−0.427 [−0.443, −0.411]	0.0	100.0
Curvature, $\kappa = 0.25$	**0.313**	0.297	0.353	0.434	0.321	+0.016 [+0.012, +0.020]	16.8	0.0
Curvature, $\kappa = 0.5$	**0.376**	0.445	0.324	0.391	0.420	−0.069 [−0.075, −0.062]	26.7	0.0
Curvature, $\kappa = 1$	**0.373**	0.650	0.275	0.324	0.579	−0.278 [−0.288, −0.268]	53.9	0.0

The registered tests are listed in Table 32. As Table 27 shows, OP beats the same-calendar trial at M0 by 0.228 [0.222, 0.234], under every tail and support-gap strength and at curvature 0.25 and 0.5. At M0, which is study E's generator without any departure, it keeps 81.5% [80.5, 82.6] of the framework's gain over that trial and is 0.134 [0.130, 0.137] below its own fallback, and it is also below the fallback under every tail strength, at support gaps 0.10 and 0.25 and at curvature 0.25. Curvature between piloted offers leaves the piloted cells unchanged, and the validation offers detect it: the flagged share rises from 12.3% at M0 to 27% and 54% at curvature 0.5 and 1, where the policy lowers regret against the framework by 0.069 [0.062, 0.075] and 0.278 [0.268, 0.288]. With half the history support removed, the reachability check sends virtually every unit to the fallback, lowering regret against the framework by 0.427 [0.411, 0.443].

G PUBLIC RETAIL DATA

M5. From M5 (Makridakis et al., 2022a) we draw 3,000 series, 300 per store, from an 873-day panel, with 45 weekly origins and seven-day targets; eligibility uses the preceding 182 days and five groups use positive-sales days in the preceding 28 days. Each forecast is trained on 13 earlier weekly origins, and 2,048 aligned scenarios keep cross-day ranks. Scale tests use all 30,490 series at three origins. Sales carry no incentive response, so M5 tests computation and settlement transfer only.

H COMPUTATION AND APPROXIMATION BOUNDARIES

Complete time runs from loading the inputs to returning every requested plan and includes all preparation; model fitting is reported separately. All methods in a comparison receive the same laws, menus, budget bands and the same permission to reuse inputs.

Reuse across requests. On 3,000 M5 series and 45 origins, reusing a valid law instead of regenerating it for each request lowers the median request time from 84.78 to 4.36 seconds, paired $20.00\times$ faster, with identical scenarios and identical exact values on all 405 tables; charging the 80.63-second preparation to the framework, 8, 32 and 127 requests per origin are $1.66\times$, $3.92\times$ and $7.11\times$ faster. On all 30,490 series the times are 82.62 and 1,146.49 seconds, a ratio of $13.88\times$, and all nine optima agree. With the 1,063.87-second preparation charged, eight requests are $1.59\times$ faster in the median over the three origins.

Per-request mixed-integer programming. The multiple-choice piecewise-linear MIP uses binaries y_s and integers t_s with

$$\sum_{s\in g} y_s = 1, \qquad 0 \le t_s \le (U_s - L_s)y_s,$$
$$x = \sum_s (L_s y_s + t_s), \qquad v = \sum_s [(a_s L_s + b_s)y_s + a_s t_s]/D,$$

solved to zero gap with a time limit per request; every answer is checked in exact arithmetic. Over 396 conditions, 180 on M5 and 216 controlled, with 33 budget bands each, the dynamic program reuses its value curve across requests and is $5.52\times$ faster than the envelope MIP at 33 requests, and raw dense allocation is $5.47\times$ faster, with the same optimal value on every request the MIP resolves. Each condition's 33 budgets were answered in five balanced method orders, a timing repetition, which gives 65,340 requests per method; the raw and envelope MIP formulations leave 25 and 5 unresolved, while all dynamic-programming variants resolve every request. The gain over per-request MIP comes from reusing the value curve.

New payment rule with the same laws. For volume tier J_{is} and active days D_{is}, the active-day rule pays $7D_{is}\rho_{J_{is}}$ when $D_{is} \ge d$ and $J_{is} > 0$. The histogram

$$H_s(k, j) = \sum_{i\in G_l} \mathbf{1}\{D_{is} = k, J_{is} = j\}$$

gives the exact coefficients

$$A_{sj}(d) = 7\sum_{k=d}^{7} kH_s(k, j)$$

for all eight attendance requirements. Over 45 origins, 2,048 scenarios, five groups and 270 reward paths, all 14,580 answers equal direct settlement, and shared preparation cuts complete time at every origin, by 70.98% on average and by 67.41–74.08% across origins.

I CONTROLLED RESPONSE INSTANCE

Support, settlement and objective. Five groups share 32 weekly trajectory labels, 29 of them distinct, with daily completions

$$y_{kd} = \begin{cases} 2(\lfloor k/8 \rfloor + 1) + ((d + k) \bmod 3), & d < k \bmod 8, \\ 0, & \text{otherwise}, \end{cases}$$

and $C_k = \sum_d y_{kd}$, $D_k = \sum_d \mathbf{1}\{y_{kd} > 0\}$. Groups use fixed-prize, per-completion, attendance, fixed-prize and per-completion rules. Each group has nine offers $a = 3b + t$, $b, t \in \{0, 1, 2\}$, with rung rates $(\ell_j + 40b + 10t)/100$, $\ell = (10, 15, 20)$. Fixed prizes and per-completion rates use completion rungs $(14, 28, 42)$ and pay $\mathrm{rung}_j\, r_j$ or $C_k r_j$; attendance pays $7D_k r_j$ at rungs of $(2, 4, 6)$ active days; below the first rung the payment is zero. With uplift U_g, mean payment M_g and stability $A_g = \Pr(0.8M_g \le Y_g \le 1.2M_g)$, the composite objective is

$$J = \sum_g [U_g + M_g A_g]/104;$$

requests are $B = 0, 5, \ldots, 160$ with band $[0.9B, 1.1B]$ on expected total payment, over the $9^5 = 59{,}049$ portfolios.

Response laws. With $x = 2b + t$, base weights $w_{gk} \in \{1, \dots, 17\}$, elasticity e_g, threshold τ_g and rule index j, offer-conditioned masses are proportional to $w_{gk} f_{gk}(x)$ with

$$\begin{aligned} f^{\text{null}} &= 1, \\ f^{\text{sat}} &= 128 + e_g(j+1)\tfrac{x}{2+x} C_k, \\ f^{\text{thr}} &= 128 + e_g(j+1)(x - \tau_g)_+^2 C_k, \\ f^{\text{trade}} &= 256 + e_g\big[(j+1)\tfrac{16x}{1+x} D_k + \tfrac{x(6-x)}{6} C_k\big]. \end{aligned}$$

The null and threshold families add a shared weekly shock z through the factor

$$32 + u_g\big[zC_k + (1-z)(C_{\max} - C_k)\big];$$

the no-offer law omits f.

J SETTLEMENT INTERFACE AND EXACT ALLOCATION

Table 28 lists the seven modules of Figure 1 under the figure's short names, their inputs and outputs, and the instance used on delivery data. Each module can be replaced by any estimator with the same input and output; allocation exactness says nothing about the accuracy of Modules 1–5.

Algorithm 1 Answering a request $r = (\mathcal{I}, W, x, \Gamma, \mathcal{B})$

1: **Once per origin:** fit M3 on pre-origin history; draw R joint per-day samples for every rider with cross-day ranks from each rider's history; store.
2: **Once per rule family:** run a pilot of n rider-weeks at three offers; fit $\lambda_{l,a}$ by Eq. (5).
3: M1–M2 and M4: select roster $\mathcal{I}$ and form layers $\mathcal{G}$ with plan menus $\mathcal{A}_l$ from recent activity.
4: **for** each layer l and plan $a \in \mathcal{A}_l$ **do**
5: M5: compose window W from stored samples; tilt to $q_{l,a}$ by reweighting the stored samples, or the trajectory types when they can be enumerated.
6: M6: compute award probabilities and $(\widehat{c}_l(a), \widehat{v}_l(a))$ for rule Γ as the weighted sum of Eq. (4) over the stored samples, the exact sum when types are enumerable; no redraw.
7: **end for**
8: M4 and M7: reduce menus by the rule admissible for $\mathcal{B}$; solve Eq. (2) on $\mathcal{B}$ by exact enumeration or the budget-grid program.
9: **return** plans with attaining rewards, or report that no plan satisfies $\mathcal{B}$.

Table 28: Four stages and seven modules. The last column gives the instance used on the delivery platform.

Stage	Module	Input → output	Delivery instance
Predict	1. Rank: response ranking	pre-decision context and past offers → ranking by incremental activity	doubly robust learner with a 5% no-offer group (Dudík et al., 2011)
	2. Recall: roster	ranking and eligibility → population fixed before attendance, incl. zero-activity riders	at least one delivery in the preceding 182 days
	3. Forecast: baseline law	histories and context → no-offer outcome law for the requested window	recurrent network, 60-day history, five-day zero-inflated Beta shares
Reduce	4. Activity layers: grouping	baseline capacity → groups and non-dominated plan menus	ordered buckets merged to balance dispersion and spending stability
Integrate	5. Correct: response correction	no-offer law and pilot or context-matched outcomes → offer-conditioned joint trajectories	exponential tilt, Eq. (5); residual matching on incentive, time and weather
	6. Score: settlement	trajectories, payment rule, attainable rewards → cost, award probability, score and attaining reward	Eqs. (6)–(7)
Allocate	7. Allocate: budget allocation	group domains and both budget bounds → reward portfolio or infeasibility	two-sided Bellman recursion, Eq. (11)

Settlement identities. For nested award events $E_1 \supseteq \cdots \supseteq E_J$ and $E_{J+1} = \varnothing$, a highest-rung prize pays

$$g = \sum_j R_j \mathbf{1}\{E_j \setminus E_{j+1}\};$$

taking expectations gives Eq. (6) without independence across riders. A rate paid on every completion satisfies, by telescoping,

$$g = \sum_j (\rho_j - \rho_{j-1}) Y \mathbf{1}\{E_j\},$$

which gives Eq. (7). For integrable quantile functions,

$$\mathbb{E}Y(a) - \mathbb{E}Y(0) = \int_0^1 [Q_a(v) - Q_0(v)]\, dv, \tag{9}$$

so uplift is integrated once on the final response curve, which avoids counting overlapping rung effects twice; the identity is an accounting statement, not causal identification. Mean payment and award probability do not determine spending stability: if one or two completions are equally likely, a fixed prize $3/2$ and a unit rate 1 have the same mean payment and award probability, but they fall in $[1.2, 1.8]$ with probability one and zero, respectively. This is why M6 integrates joint trajectories.

J.1 Exact allocation over attaining score segments

Fix a conditional trajectory law and uplift u_{lq} on each group/path (l, q). On a straight reward path $r(t) = r^- + t(r^+ - r^-)$, $0 \le t \le 1$, with expected-payment endpoints $\mu^- < \mu^+$,

$$\begin{aligned} t(\mu) &= \frac{\mu - \mu^-}{\mu^+ - \mu^-}, \\ S_s(\mu) &= k_s^\top r(t(\mu)), \end{aligned} \tag{10}$$

where k_s is the payment coefficient vector of scenario s under the rule. Each closed event $0.8\mu \le S_s(\mu) \le 1.2\mu$ is an interval, a point or empty, so its probability $p_{lq}(\mu)$ is piecewise constant and the score

$$v_{lq}(\mu) = u_{lq}/T_0 + \lambda \mu p_{lq}(\mu)/T_0$$

is piecewise affine, with T_0 a common normalizer. Gaps between paths are kept, never interpolated, and a path is used only inside one response branch.

Same-spend reduction. For attainable sets I_{lq} keep the envelope $F_l(b) = \max_{q: b \in I_{lq}} v_{lq}(b)$ with an attaining reward. With additive scores this preserves the optimum under any constraint that depends only on the group amounts: each candidate can be replaced by a no-worse attaining plan at the same amount, and every such plan was feasible. Lemma 1 states which menu reductions are admissible for which budget.

Lemma 1. Admissible menu reductions. *Let each layer l choose one plan from a finite menu $\mathcal{A}_l$ with cost $c_l(a) \ge 0$ and value $v_l(a)$, and let a portfolio be feasible when $\sum_l c_l(a_l) \in \mathcal{B}$. (i) Under a sole cap $\mathcal{B} = [0, \bar{b}]$, deleting a plan a for which some $a' \in \mathcal{A}_l$ has $c_l(a') \le c_l(a)$ and $v_l(a') \ge v_l(a)$, with ties broken by a fixed order, preserves feasibility and the optimal value. (ii) For any $\mathcal{B}$, deleting a when some a' has $c_l(a') = c_l(a)$ and $v_l(a') \ge v_l(a)$ preserves feasibility and the optimal value; in both cases the retained plan takes the place of the deleted one in an optimal portfolio. (iii) Rule (i) is not admissible once $\mathcal{B}$ has a positive floor.*

Proof. (i) Replacing a_l by a' in a feasible portfolio lowers total cost weakly, so it stays in $[0, \bar{b}]$, and raises value weakly; repeating the replacement ends on retained plans, since the fixed tie order excludes cycles. (ii) The replacement leaves total cost unchanged, so membership in any $\mathcal{B}$ is unchanged. (iii) One layer with plans $(c, v) = (1, 2), (2, 1)$ and $\mathcal{B} = [1.8, 2.2]$: rule (i) deletes the second plan, which is the only feasible one. □

Which rule each experiment uses, and an independent check. Rider-record budgets are sole caps, so M4 applies rule (i) and the exact reference enumerates the products of the rule-(i) menus. The controlled study of Section 5 has two-sided bands and applies no reduction: it enumerates all 9^5 portfolios. The segment and envelope allocators below use only rule (ii). To rule out an error that a reduction shared by solver and reference would hide, we re-solved every rider table small enough for complete enumeration, with at most 2×10^8 portfolios, which covers 675 of the 810 realized and 12,150 of the 14,580 predicted tables, by a separately written complete enumeration without any reduction: feasibility and optimal value agree with the reference on every one of them; only the

returned plan can differ among portfolios of equal value. The budget-grid program loses value on 3 of the 405 predicted five-layer main-grid instances, by at most 0.92%, and never exceeds the cap. In the controlled study, 10,800 enumeration calls covering 356,400 budget bands differ in 0, and rerunning the study with the independent solver's choices reproduces every stored result exactly.

Segment recursion. Let $\delta = T_0/h$, $o_l = \min_q \mu^-_{lq}/\delta$ and $\mu_l = \delta(x_l + o_l)$ with integer x_l. With R equally weighted scenarios and stability count n, candidate q of group l attains score $\alpha x + \beta$, with

$$\alpha = \lambda n/(Rh),$$
$$\beta = u_{lq}/T_0 + \alpha o_l,$$

at every integer spend index $x \in [L, U]$ of a segment. A physical band $[B^-, B^+]$ becomes the integer interval

$$\Big[\max\big\{0, \lceil B^-/\delta - \textstyle\sum_l o_l\rceil\big\},\ \lfloor B^+/\delta - \sum_l o_l\rfloor\Big].$$

Each segment contributes

$$D_l(b) = \max_{q,s}\Big\{\alpha_{qs} b + \beta_{qs} + \max_{k\in[b-U_{qs},\, b-L_{qs}],\ 0\le k<N}\big[D_{l-1}(k) - \alpha_{qs} k\big]\Big\}. \tag{11}$$

Minimum-cost options, closed endpoints, gaps and unreachable states are retained, and no no-offer action is added.

Proposition 2. Segment and dense allocation coincide. *For a fixed law, additive scores, group-wise menus and exact rational arithmetic, the recursion* (11) *returns the same value curve and the same canonical rewards as dense allocation on the same grid.*

Proof. Step 1: the recursion. Set $D_0(0) = 0$ and all unreachable values to $-\infty$. On a segment, substitute $x = b - k$ in $D_{l-1}(b - x) + \alpha x + \beta$; the feasible predecessor interval gives Eq. (11), and maximizing over segments keeps every action.

Step 2: ties. Ties within a segment keep the largest k, that is the smallest current x; ties across segments keep the smallest x, then the smallest candidate index. By induction the choices equal the dense ones.

Step 3: rewards. Backtracking from the smallest optimal feasible total through Eq. (10) returns rewards with the same score and physical spend; an unreachable band is infeasible for both. □

A monotone queue evaluates each window maximum in $O(N)$, so K segments, L groups and N budget states need $O(KN + LK + LN)$ operations against $O(LN^2 + KN + LK)$ for dense allocation; both store $O(LN)$ choices. If each group's candidate intervals are pairwise disjoint, the envelope equals the raw menu and its construction can be skipped without changing the solution. Compared with graphical dynamic programming for piecewise-linear investment under an upper cap (Gafarov et al., 2014; 2016) and nonconvex piecewise-linear knapsacks (Kameshwaran & Narahari, 2009), the recursion handles reward domains with gaps, two-sided budgets, rule-dependent event scores and an attaining reward for every scored spend.

K FOUR-TERM LOSS DECOMPOSITION: PROOFS AND STAGE BOUNDS

Setting. Layers $l = 1, \dots, L$ have finite plan menus $\mathcal{A}_l$; a portfolio is $a = (a_1, \dots, a_L)$ with

$$c(a) = \sum_l c_l(a_l),$$
$$v(a) = \sum_l v_l(a_l).$$

Tables without a superscript are the true ones; $(c^{(k)}, v^{(k)})$ for $k = \mathrm{syn}, \mathrm{mm}, \mathrm{int}, \mathrm{dec}$ are the successive tables of the framework and $(\widehat{c}, \widehat{v}) = (c^{\mathrm{dec}}, v^{\mathrm{dec}})$. Stage errors $\varepsilon^c_k, \varepsilon^v_k$ and their sums E_c, E_v are

as in Section 3. The band and the value it carries are

$$\begin{aligned}\mathcal{B} &= [\underline{b}, \bar{b}],\\ \mathcal{B}_\delta &= [\underline{b} + \delta, \bar{b} - \delta],\\ V^\star(\delta) &= \max\{v(a) : c(a) \in \mathcal{B}_\delta\},\\ \omega(\delta) &= V^\star(0) - V^\star(\delta) \geq 0,\end{aligned}$$

with $V^\star(\delta) = -\infty$ if no portfolio qualifies. Cost and uplift entries are expectations of bounded functions of the trajectory such as payment and completions, or differences of two such expectations, so they are linear in the law; the same holds for a stability indicator with a fixed band. Parts (i)–(ii) use only table errors and therefore also cover the mean-centred stability score, which is not linear in the law. In particular, if a layer law is the mixture of its members' laws, n_l times its expectation equals the sum of the members' expectations exactly; population reduction therefore adds no error to expected tables and acts only through the menus, which enter $V^\star$ and ω.

K.1 Proof of Proposition 1

(i) Upper bound. *Step 1: table errors.* For every portfolio, the triangle inequality along the chain gives

$$\begin{aligned}|c(a) - \widehat{c}(a)| &\leq \sum_l \max_{a' \in \mathcal{A}_l} |c_l(a') - \widehat{c}_l(a')|\\ &\leq \sum_l \sum_k \max_{a'} |c_l^{(k)}(a') - c_l^{(k^-)}(a')|\\ &= E_c,\end{aligned}$$

and likewise $|v(a) - \widehat{v}(a)| \leq E_v$.

Step 2: feasibility. Since $\widehat{c}(\widehat{a}) \in \mathcal{B}_{E_c}$,

$$c(\widehat{a}) \in [\widehat{c}(\widehat{a}) - E_c,\ \widehat{c}(\widehat{a}) + E_c] \subseteq \mathcal{B}.$$

Step 3: value. Let a° attain $V^\star(2E_c)$. Then $\widehat{c}(a^\circ) \in \mathcal{B}_{E_c}$, so a° is admissible for the planner, the admissible set is nonempty, and $\widehat{v}(\widehat{a}) \geq \widehat{v}(a^\circ) - \eta$. Hence

$$\begin{aligned}v(\widehat{a}) &\geq \widehat{v}(\widehat{a}) - E_v && \text{by Step 1}\\ &\geq \widehat{v}(a^\circ) - \eta - E_v && \text{by } \eta\text{-optimality}\\ &\geq v(a^\circ) - 2E_v - \eta && \text{by Step 1}\\ &= V^\star(2E_c) - 2E_v - \eta.\end{aligned}$$

Step 4: conclusion. Subtracting from $V^\star(0)$ gives Eq. (8). If $\omega(\delta) \leq \kappa\delta$ on $[0, 2E_c]$,

$$\omega(2E_c) \leq 2\kappa E_c = \sum_k 2\kappa\varepsilon_k^c.$$

If the comparator ranges over a finer reward domain than the planner's menu, the difference between the two optima at $2E_c$ is added to η. The proof uses only $|c(a) - \widehat{c}(a)| \leq E_c$ and $|v(a) - \widehat{v}(a)| \leq E_v$, so (i) also holds with the direct errors

$$\begin{aligned}\bar{E}_c &= \sum_l \max_a |c_l(a) - \widehat{c}_l(a)| \leq E_c,\\ \bar{E}_v &\leq E_v\end{aligned}$$

in their place. □

(ii) Lower bound. *Step 1: the instance.* Fix stage errors $(\varepsilon_k^c, \varepsilon_k^v)_k$ with $E_c > 0$, leaving the case $E_c = 0$ to Step 5, any $\Delta > 0$, and a band with $\bar{b} - \underline{b} > 4E_c$ and midpoint m. Order the stages as syn, mm, int, dec and write $s_k(\varepsilon) = \sum_{j \le k} \varepsilon_j$ for the partial sums along this order. Layer 1 has plans s and h with true $(c, v) = (m, 0)$ and $(\bar{b} - E_c, \Delta)$. Plan s is exact at every stage; plan h has exact value and cost

$$c^{(k)}(h) = \bar{b} - E_c + s_k(\varepsilon^c),$$

so that $\widehat{c}(h) = \bar{b}$. Layer 2 has plans x, y of zero cost at every stage, true values 0 and $2E_v$, and

$$v^{(k)}(x) = s_k(\varepsilon^v),$$
$$v^{(k)}(y) = 2E_v - s_k(\varepsilon^v),$$

so that $\widehat{v}(x) = \widehat{v}(y) = E_v$. Each stage error is exactly the prescribed one.

Step 2: the benchmark. In $\mathcal{B}_{2E_c}$ plan h is excluded. For this truth

$$V^\star(0) = \Delta + 2E_v \qquad \text{attained by } (h, y) \text{ at spend } \bar{b} - E_c \in \mathcal{B},$$
$$V^\star(2E_c) = 2E_v \qquad \text{attained by } (s, y),$$
$$\omega(2E_c) = \Delta.$$

Step 3: equality for the planner of (i). With $\eta = 0$ it cannot use h because $\widehat{c}(h) = \bar{b} \notin \mathcal{B}_{E_c}$; breaking the tie in layer 2 toward x it earns 0, so Eq. (8) holds with equality.

Step 4: every planner. Now take any deterministic planner that sees the final tables and the stage-error magnitudes and keeps the true spend in $\mathcal{B}$ for every consistent truth. The truth with $c(h) = \bar{b} + E_c$, which has the same final tables and is reached by decreasing stage steps, makes every portfolio with h infeasible, so the planner never selects h. In layer 2 its choice is a function of identical final tables, so the truth that gives value 0 to the chosen plan and $2E_v$ to the other is consistent; the planner earns 0 against $V^\star(0) = \omega(2E_c) + 2E_v$. Abstaining also earns 0. A randomized planner loses at least $\omega(2E_c) + E_v$ in expectation, by averaging over the two value truths. Without tightening, a planner that maximizes $\widehat{v}$ over $\widehat{c} \in \mathcal{B}$ selects $(h, \cdot)$ with estimated spend $\bar{b}$, whose true spend is $\bar{b} + E_c$ under the second truth. Adding a layer with two exact zero-cost plans of value 0 and η, an η-optimal planner may choose the first, so equality also holds for $\eta > 0$.

Step 5: the case $E_c = 0$. Then $\mathcal{B}_{E_c} = \mathcal{B}$, plan h would stay admissible and $\omega(0) = 0$, so drop h: layer 1 has only the exact plan s and layer 2 is as above. Now every cost is exact and $V^\star(0) = 2E_v$, attained via (s, y) at spend $m \in \mathcal{B}$. The planner of (i) breaking the tie toward x earns 0, and by the argument for layer 2 any deterministic planner earns 0 against one of the two value truths. Hence the bound

$$\omega(0) + 2E_v = 2E_v$$

is attained. □

(iii) Error floors of each omitted stage. All instances use exact tables for the stages that are kept.

Pilot correction. Take one layer with two offers whose payment functions coincide, true offer laws P_{a_1} equal to the no-offer law and P_{a_2} with expected completions larger by $U > 0$, and a band containing both costs. Without the correction both synthetic laws equal the no-offer law, the two offers have identical tables and estimated uplift 0, and relabelling the offers makes any deterministic planner choose a_1: it loses the entire optimal uplift U.

History shape. By Lemma 2 two laws with the same mean and variance can differ by $1/(1 + t^2)$ in the probability of reaching $\mu + t\sigma$, so any estimate computed from the two moments alone errs by at least $1/(2(1 + t^2))$ on one of them, and a fixed prize R at that threshold is mispriced by at least $R/(2(1 + t^2))$.

Trajectory integration. Let $Y \in \{0, 2\tau\}$ with probability $\frac{1}{2}$ each, known exactly, and a prize R paid when $Y \ge \tau$. Evaluating the rule at the mean gives

$$R\mathbf{1}\{\mathbb{E}Y \ge \tau\} = R,$$

whereas

$$\mathbb{E}[R\mathbf{1}\{Y \ge \tau\}] = R/2.$$

The probability that payment lies within $\pm 20\%$ of its mean $R/2$ is 0, while the plug-in value is 1.

Joint allocation. Under the cap $\mathcal{B} = [0, B]$ with $L \geq 2$ layers, let layer 1 offer $(0, 0)$ and (B, U) and all other layers offer only $(0, 0)$. The joint optimum earns U; splitting the budget equally gives layer 1 at most $B/L < B$, so it earns 0. □

Lemma 2. Two moments do not fix a threshold probability. *For $t > 0$,*

$$\sup\left|\Pr(S \geq \mu + t\sigma) - \Pr(S' \geq \mu + t\sigma)\right| = \frac{1}{1 + t^2}$$

over pairs of laws with common mean μ and variance $\sigma^2 > 0$, and the supremum is attained. For integer laws, $\{0{:}\frac{1}{2}, 2{:}\frac{1}{2}\}$ and $\{0{:}\frac{1}{3}, 1{:}\frac{1}{2}, 3{:}\frac{1}{6}\}$ both have mean and variance 1 but $\Pr(S \geq 2)$ equal to $\frac{1}{2}$ and $\frac{1}{6}$.

Proof. Upper bound. For $u > 0$,

$$\Pr(S - \mu \geq t\sigma) \leq \frac{\mathbb{E}[(S - \mu + u)^2]}{(t\sigma + u)^2} = \frac{\sigma^2 + u^2}{(t\sigma + u)^2},$$

which equals $1/(1 + t^2)$ at $u = \sigma/t$; both probabilities lie in $[0, 1/(1 + t^2)]$.

Equality. $S = \mu + t\sigma$ with probability $1/(1 + t^2)$ and $S = \mu - \sigma/t$ otherwise has mean μ and variance σ^2; S' on $\{\mu + \alpha, \mu - \sigma^2/\alpha\}$ with $0 < \alpha < t\sigma$ and mass $\sigma^2/(\alpha^2 + \sigma^2)$ at $\mu + \alpha$ has the same moments and never reaches $\mu + t\sigma$. The integer example is checked directly. □

K.2 Bounds on the four stage terms

Lemma 3. Award probabilities carry every threshold rule. *Let $Y \in \{0, \ldots, M\}$ count eligible completions, D be a common attendance or eligibility event, $E_j = \{Y \geq \tau_j\} \cap D$ with integers $1 \leq \tau_1 < \cdots < \tau_J$, and $G_F(k) = F(Y \geq k, D)$. For laws F, F' write $\Delta(k) = |G_F(k) - G_{F'}(k)|$. For fixed prizes $0 = R_0 \leq R_1 \leq \cdots \leq R_J$ paid at the highest rung reached,*

$$|\mathbb{E}_F g - \mathbb{E}_{F'} g| \leq \sum_j (R_j - R_{j-1})\Delta(\tau_j) \leq R_J \max_j \Delta(\tau_j).$$

For rates $0 = \rho_0 \leq \cdots \leq \rho_J$ paid on all eligible completions at the highest rung reached,

$$\begin{aligned}|\mathbb{E}_F g - \mathbb{E}_{F'} g| &\leq \sum_j (\rho_j - \rho_{j-1})\Big\{\tau_j \Delta(\tau_j) + \sum_{k > \tau_j} \Delta(k)\Big\} \\ &\leq \rho_J M \max_{k \geq \tau_1} \Delta(k).\end{aligned}$$

Expected completions satisfy

$$|\mathbb{E}_F Y - \mathbb{E}_{F'} Y| \leq \sum_{k=1}^{M} |F(Y \geq k) - F'(Y \geq k)|.$$

Proof. Nested events give

$$g = \sum_j (R_j - R_{j-1})\mathbf{1}\{E_j\} \qquad \text{for fixed prizes,}$$

$$g = \sum_j (\rho_j - \rho_{j-1}) Y \mathbf{1}\{E_j\} \qquad \text{for rates.}$$

With $p(y) = F(Y = y, D)$, summation by parts gives

$$\begin{aligned}\mathbb{E}_F[Y\mathbf{1}\{E_j\}] &= \sum_{y \geq \tau_j} y p(y) \\ &= \tau_j G_F(\tau_j) + \sum_{k > \tau_j} G_F(k),\end{aligned}$$

and, in the same way,

$$\mathbb{E}_F Y = \sum_{k \geq 1} F(Y \geq k).$$

Take differences; at most $M - \tau_j$ terms follow τ_j. □

Thus each table error of a threshold rule is controlled by errors in award probabilities $G(k)$. When D requires attendance on several days, G depends on the joint law of completions and attendance, not on their separate marginals: over two days with $D = \{\text{both days active}\}$ and $\tau_1 = 3$, the laws $\frac{1}{2}\delta_{(2,1)} + \frac{1}{2}\delta_{(3,2)}$ and the uniform law on $\{(2,1),(2,2),(3,1),(3,2)\}$ of completions and active days share both marginals but give $G(3) = \frac{1}{2}$ and $\frac{1}{4}$. This is why integration uses joint trajectories.

Lemma 4. Synthetic term. *On a finite trajectory space let p_0 be the no-offer law, ϕ the matched features, P a true offer law with $P \ll p_0$, and $Q_\lambda \propto p_0 e^{\lambda^\top \phi}$; KL denotes the Kullback–Leibler divergence and TV the total variation distance. If $\lambda^\star$ solves $\mathbb{E}_{Q_{\lambda^\star}} \phi = \mathbb{E}_P \phi$, then*

$$\mathrm{KL}(P\|p_0) = \mathrm{KL}(P\|Q_{\lambda^\star}) + \mathrm{KL}(Q_{\lambda^\star}\|p_0),$$

and for every function f with range $\mathrm{osc}(f) = \max f - \min f$,

$$|\mathbb{E}_P f - \mathbb{E}_{Q_{\lambda^\star}} f| \;\leq\; \mathrm{osc}(f)\sqrt{\tfrac{1}{2}\{\mathrm{KL}(P\|p_0) - \mathrm{KL}(Q_{\lambda^\star}\|p_0)\}}.$$

Proof. Step 1: the identity.

$$\begin{aligned}\mathrm{KL}(P\|p_0) - \mathrm{KL}(P\|Q_{\lambda^\star}) &= \mathbb{E}_P \log(Q_{\lambda^\star}/p_0) \\ &= \lambda^{\star\top}\mathbb{E}_P \phi - \log Z,\end{aligned}$$

and $\mathrm{KL}(Q_{\lambda^\star}\|p_0) = \lambda^{\star\top}\mathbb{E}_{Q_{\lambda^\star}}\phi - \log Z$ is the same number.

Step 2: the bound. Then $|\mathbb{E}_P f - \mathbb{E}_Q f| \leq \mathrm{osc}(f)\,\mathrm{TV}(P,Q)$ and Pinsker's inequality $\mathrm{TV} \leq \sqrt{\mathrm{KL}/2}$ (Tsybakov, 2009, Lemma 2.5). □

The same identity applied to any Q with the matched moments shows $\mathrm{KL}(Q\|p_0) \geq \mathrm{KL}(Q_{\lambda^\star}\|p_0)$: the tilt is the least change of history that reproduces the experiment (Csiszár, 1975). Without the correction the same argument gives only $\mathrm{osc}(f)\sqrt{\mathrm{KL}(P\|p_0)/2}$; matching removes $\mathrm{KL}(Q_{\lambda^\star}\|p_0)$ from the radicand, and the term vanishes when the response is a tilt in ϕ. The bound covers piloted offers. At an unpiloted offer, P^{syn} uses the interpolated parameter

$$\lambda_t^\star = (1-t)\lambda_j^\star + t\lambda_k^\star$$

of its piloted neighbours j, k as in Section 3, so the synthetic term also contains the interpolation bias $|\mathbb{E}_P f - \mathbb{E}_{Q_{\lambda_t^\star}} f|$, which the lemma does not bound and which persists however large the pilot. For example, with $Y \sim \text{Bernoulli}(\frac{1}{2})$ under no offer, $\phi(Y) = Y$ and a true response $P_a(Y=1) = \sigma(a^2)$ inside the tilt family, with σ the logistic function, pilots at $a = 0.2$ and 0.6 give $\lambda^\star = 0.04$ and 0.36 exactly, but the interpolated $\lambda = 0.20$ at $a = 0.4$ differs from the true 0.16, leaving an error of $\sigma(0.20) - \sigma(0.16) \approx 0.0099$ in $\mathbb{E}Y$ with exact moments and exact integration. More pilot offers, or smoothness of λ in the offer, reduce this part.

Lemma 5. Moment-matching term. *Keep p_0 fixed and let $\Sigma_\lambda = \mathrm{Cov}_{Q_\lambda}(\phi)$ satisfy $\kappa_\phi I \preceq \Sigma_\lambda \preceq \Lambda_\phi I$, $\kappa_\phi > 0$, on the segment between $\lambda^\star$ and the fitted $\widehat{\lambda}$, whose moments $\widehat{m}$ are the pilot means. Then for every f,*

$$\begin{aligned}|\mathbb{E}_{Q_{\widehat{\lambda}}} f - \mathbb{E}_{Q_{\lambda^\star}} f| &\leq \tfrac{1}{2}\mathrm{osc}(f)\,\Lambda_\phi^{1/2}\,\|\widehat{\lambda} - \lambda^\star\|_2 \\ &\leq \tfrac{1}{2}\mathrm{osc}(f)\,\frac{\Lambda_\phi^{1/2}}{\kappa_\phi}\,\|\widehat{m} - m^\star\|_2.\end{aligned}$$

If each of J_p piloted offers per layer receives n independent trajectories and $\phi \in [0,1]^d$, then with probability at least $1-\delta$, simultaneously for all layers and piloted offers,

$$\|\widehat{m} - m^\star\|_2 \leq \sqrt{d\log(2dLJ_p/\delta)/(2n)}.$$

Offers whose λ is interpolated linearly between piloted offers, or between $\lambda = 0$ at the no-offer plan and a piloted offer, satisfy the first inequality, given the covariance bounds on their own segment, with $\lambda^\star$ the interpolated synthetic parameter and $\|\widehat{\lambda} - \lambda^\star\|_2$ at most the larger error of their two neighbours; the comparison is with P^{syn}, not with the projection of the offer's own population moments.

Proof. Step 1: derivative along the segment. With $u = \widehat{\lambda} - \lambda^\star$ and $\lambda_t = \lambda^\star + tu$,

$$\begin{aligned}\Big|\frac{d}{dt}\mathbb{E}_{Q_{\lambda_t}} f\Big| &= |\mathrm{Cov}_{\lambda_t}(f, u^\top \phi)| \\ &\le \mathrm{sd}(f)\,(u^\top \Sigma_{\lambda_t} u)^{1/2} \\ &\le \tfrac{1}{2}\mathrm{osc}(f)\Lambda_\phi^{1/2}\|u\|.\end{aligned}$$

Step 2: from parameters to moments. Also

$$\widehat{m} - m^\star = \int_0^1 \Sigma_{\lambda_t} dt\, u$$

with

$$\int_0^1 \Sigma_{\lambda_t} dt \succeq \kappa_\phi I,$$

so $\|u\| \le \|\widehat{m} - m^\star\|/\kappa_\phi$.

Step 3: the rate. Hoeffding's inequality (Hoeffding, 1963) for each of the dLJ_p coordinates with a union bound gives the rate.

Step 4: interpolated offers. Interpolation weights are convex, so the interpolated error is a convex combination of the neighbours' errors. □

If the no-offer law is estimated from N_0 history weeks, changing the base at fixed λ moves every expectation by at most

$$2\,\mathrm{osc}(f) e^{s(\lambda)} \mathrm{TV}(\widehat{p}_0, p_0),$$

where

$$s(\lambda) = \max_z \lambda^\top \phi(z) - \min_z \lambda^\top \phi(z),$$

since, for $w = e^{\lambda^\top \phi}$,

$$\|Q - Q'\|_1 \le 2 \sum_z w|p_0 - \widehat{p}_0|/\mathbb{E}_{p_0} w;$$

and $\mathbb{E}\,\mathrm{TV}(\widehat{p}_0, p_0) \le \frac{1}{2}\sqrt{|\mathcal{Z}|/N_0}$. With abundant history, $N_0 \gg n$, this part is small; the measured moment term includes it.

This bound holds at fixed λ. The method, however, re-fits the tilt on $\widehat{p}_0$ so that it matches the pilot moments again, and the change of λ must be controlled as well.

Lemma 6. Estimated base and re-fitted tilt. *Let $Q = Q_\lambda[p_0]$ be the tilt of p_0 with moments m, let $Q' = Q_\lambda[\widehat{p}_0]$ keep λ on the estimated base, and let $\widehat{Q} = Q_{\widehat{\lambda}}[\widehat{p}_0]$ be re-fitted to the same moments m. Assume support overlap, that is, m lies in the interior of the convex hull of $\phi(\mathrm{supp}\,\widehat{p}_0)$. Let the feature covariance of the tilts of $\widehat{p}_0$ satisfy $\widehat{\kappa} I \preceq \Sigma \preceq \widehat{\Lambda} I$, $\widehat{\kappa} > 0$, on the segment between λ and $\widehat{\lambda}$. Put*

$$\tau = 2e^{s(\lambda)}\mathrm{TV}(\widehat{p}_0, p_0), \qquad D_\phi = \Big(\sum_i \mathrm{osc}(\phi_i)^2\Big)^{1/2}.$$

Then for every f,

$$|\mathbb{E}_{\widehat{Q}} f - \mathbb{E}_Q f| \;\le\; \mathrm{osc}(f)\,\tau\Big(1 + \tfrac{1}{2} D_\phi \frac{\widehat{\Lambda}^{1/2}}{\widehat{\kappa}}\Big).$$

Proof. Step 1: base change at fixed λ. By the display above, $\mathrm{TV}(Q', Q) \le \tau$, hence $|\mathbb{E}_{Q'} f - \mathbb{E}_Q f| \le \mathrm{osc}(f)\tau$.

Step 2: moment shift. Let m' be the moments of Q'. Each coordinate moves by at most $\mathrm{osc}(\phi_i)\tau$, so $\|m' - m\|_2 \le D_\phi \tau$.

Step 3: re-fit on the estimated base. Q' and $\widehat{Q}$ are tilts of the same base $\widehat{p}_0$ with moments m' and m. Lemma 5, whose proof uses only a fixed base and the covariance bounds on the segment, gives

$$\begin{aligned} |\mathbb{E}_{\widehat{Q}} f - \mathbb{E}_{Q'} f| &\le \tfrac{1}{2}\mathrm{osc}(f)\, \frac{\widehat{\Lambda}^{1/2}}{\widehat{\kappa}}\, \|m - m'\|_2 \\ &\le \tfrac{1}{2}\mathrm{osc}(f)\, \frac{\widehat{\Lambda}^{1/2}}{\widehat{\kappa}}\, D_\phi \tau. \end{aligned}$$

The triangle inequality concludes. □

The diagnostics of Appendix E are proxies for the three conditions, not a certification of them: support overlap, meaning that the target lies inside the range of the stored weeks, with the exact reachability test of study G in Appendix E deciding whether it lies in the interior of their hull; the weight range $e^{s(\lambda)}$, which bounds $N \max_i w_i$ on N equally weighted stored weeks; and the conditioning $\widehat{\Lambda}/\widehat{\kappa}$ of the feature covariance. If the estimated base loses support on which the matched moments rely, $\widehat{\kappa}$ becomes small or $\widehat{\lambda}$ fails to exist, and the bound grows accordingly; stability is claimed only where these quantities are controlled. The rates of Corollary 1 are stated for a given base; on the event that the conditions hold, the estimated base adds this term. When the corrected law is the tilt of the empirical stored-week law, as in Appendices D and E, the weighted sum over the stored weeks is exact for that law, and the finite history enters through $\widehat{p}_0$ in this lemma rather than through the integration term.

Experiment length. Lemma 5 needs the experiment to fix d moments at J_p offers per layer, so its error scales as $\sqrt{d/n}$. A trial that estimates the complete law of every offer directly avoids the synthetic term, but the analogous bound for empirical frequencies is $\frac{1}{2}\sqrt{|\mathcal{Z}|/n_{\mathrm{R}}}$ per offer and must hold at all $|\mathcal{A}_l|$ offers. Comparing these sufficient conditions for estimating the complete law, the trial needs roughly

$$\frac{|\mathcal{A}_l||\mathcal{Z}|}{J_p d}$$

times as many rider-weeks for the same error in total variation, up to logarithms and $\Lambda_\phi^{1/2}/\kappa_\phi$; this factor is 48 in the controlled design with $|\mathcal{Z}| = 32$, $|\mathcal{A}_l| = 9$, $J_p = 3$ and $d = 2$. The price is the synthetic term, which does not shrink with experiment length; this is a comparison of upper bounds, not a lower bound for trials. A trial that estimates only the expected payment and value of each plan on a fixed menu needs no factor $|\mathcal{Z}|$: Hoeffding's inequality with a union bound over the N layer–plan pairs gives $\mathrm{osc}(f)\sqrt{\log(4N/\delta)/(2n_{\mathrm{R}})}$ per entry, as in Lemma 7. The ratio therefore compares estimating a complete law with estimating d moments, not the trial length needed for the same decision accuracy, which Section 5 and Appendix E measure directly.

Lemma 7. Integration term. *If integration averages S independent draws from the moment-matched law for each of the $N = \sum_l |\mathcal{A}_l|$ layer–plan pairs, then with probability at least $1 - \delta$ every cost and value entry is within $\mathrm{osc}_l\sqrt{\log(4N/\delta)/(2S)}$ of its exact value, where osc_l is the range of the integrand for layer l; hence*

$$\varepsilon^c_{\mathrm{int}},\ \varepsilon^v_{\mathrm{int}} \ \le\ \sum_l \mathrm{osc}_l \sqrt{\log(4N/\delta)/(2S)}.$$

Summing exactly over a finite trajectory space gives $\varepsilon_{\mathrm{int}} = 0$.

Proof. Hoeffding's inequality for each of the $2N$ entries and a union bound. Draws may be shared across plans as common random numbers: the bound is per entry and needs no independence across entries. □

Scope of the rates. Lemmas 4–7 concern fixed integrands; Appendix K.3 gives the rates in expectation, including the conditions under which they extend to the mean-centred stability score and examples showing that these conditions are needed. The controlled experiment uses pure uplift as its primary endpoint for this reason.

Lemma 8. Decision term. *A budget-grid program with step h that rounds each layer cost to the nearest grid point and optimizes exactly over the rounded tables is an exact planner, with $\eta = 0$, on tables with $\varepsilon^{c}_{\mathrm{dec}} \le Lh/2$ and $\varepsilon^{v}_{\mathrm{dec}} = 0$. Rounding increments above each layer's minimum upward gives $0 \le c^{\mathrm{dec}} - c^{\mathrm{int}} < h$ per layer, so $\varepsilon^{c}_{\mathrm{dec}} \le Lh$; exact enumeration and the segment program on its declared monetary grid have $\varepsilon_{\mathrm{dec}} = 0$.*

Proof. Rounding moves each layer cost by at most $h/2$, or by less than h with upward rounding, and leaves values unchanged; the program maximizes exactly over the rounded problem. □

Cross-day dependence. Suppose a member's day-activity indicators are conditionally independent Bernoulli$(p(\xi))$ given a shared factor ξ over H days, so A counts active days, and let $A^{\mathrm{ind}} \sim \mathrm{Bin}(H, \mathbb{E}p(\xi))$ be the independent-day law with the same daily marginals. Then, by Jensen's inequality,

$$\Pr(A = H) = \mathbb{E}p(\xi)^H \ge (\mathbb{E}p(\xi))^H = \Pr(A^{\mathrm{ind}} = H),$$

strictly unless $p(\xi)$ is constant. Since

$$\mathbb{E}A = \mathbb{E}A^{\mathrm{ind}} = \sum_{d \ge 1} \Pr(A \ge d),$$

the two survival functions must cross: composing days independently understates the full-attendance requirement and overstates some lower one.

K.3 Rates of the moment-matching and integration terms

The rates quoted in Section 3 hold in expectation for every table entry with a fixed integrand. This subsection states them with the conditions they need, including the conditions under which they extend to a stability score whose band is centred at the law's own mean payment; the two examples below show that these conditions are needed, since without them such an entry keeps an error of constant size while the pilot moments converge.

Setting. Fix a layer and a piloted offer. The trajectory space $\mathcal{Z}$ is finite, the base p_0 has full support, and the features $\phi : \mathcal{Z} \to [0,1]^d$ do not all lie in one hyperplane. Write

$$\begin{aligned} Q_\lambda &\propto p_0 e^{\lambda^\top \phi}, \\ \Sigma_\lambda &= \mathrm{Cov}_{Q_\lambda}(\phi), \\ m(\lambda) &= \mathbb{E}_{Q_\lambda}\phi. \end{aligned}$$

The pilot mean $\widehat{m}$ averages ϕ over n independent trajectories drawn from the true offer law P; when riders of one week share a shock, n counts independent weeks or blocks, not rider-weeks. With $m^\star = \mathbb{E}_P\phi$,

$$\begin{aligned} P^{\mathrm{syn}} &= Q_{\lambda(m^\star)}, \\ P^{\mathrm{mm}} &= Q_{\lambda(\widehat{m})}; \end{aligned}$$

the base is held fixed, and its estimation adds the term stated after Lemma 5. The closed ball $\bar{B}(m^\star, \rho)$ lies in the interior of $\mathrm{conv}\,\phi(\mathcal{Z})$, and κ_ρ is the smallest eigenvalue of $\Sigma_{\lambda(m)}$ over that ball. If $\widehat{m}$ is not interior, the fit is replaced by any law or by abstention, and the entry stays in its range.

The map $\lambda \mapsto m(\lambda)$ is a bijection from $\mathbb{R}^d$ onto the interior of $\mathrm{conv}\,\phi(\mathcal{Z})$ with a smooth inverse, so $\lambda(m)$ is well defined and $\kappa_\rho > 0$. Indeed, $u^\top \Sigma_\lambda u = \mathrm{Var}_{Q_\lambda}(u^\top \phi) > 0$ for $u \ne 0$. For interior m, the strictly convex function

$$\psi(\lambda) = \log \sum_z p_0(z) e^{\lambda^\top \phi(z)} - \lambda^\top m$$

satisfies, for every unit vector u,

$$\psi(tu) \geq t\{\max_z u^\top \phi(z) - u^\top m\} + \log \min_z p_0(z) \to \infty,$$

so it has a unique minimizer, where $m(\lambda) = m$. The inverse function theorem gives smoothness, and continuity on the compact ball gives $\kappa_\rho > 0$.

Mean-centred stability. For a law q and a payment $y : \mathcal{Z} \to [0, \bar{y}]$ let

$$M(q) = \mathbb{E}_q y,$$
$$S(q) = \{z : \alpha M(q) \leq y(z) \leq \beta M(q)\}$$

with $0 < \alpha < 1 < \beta$; the paper uses 0.8 and 1.2. The monetary stability score is $F(q) = M(q)\, q(S(q))$, and its *margin*

$$\gamma(q) = \min_z \min\{|y(z) - \alpha M(q)|, |y(z) - \beta M(q)|\}$$

is the distance from the band's endpoints to the nearest payment value.

Corollary 1. Rates of the moment-matching and integration terms. *In the setting above let $M^\star = M(P^{\mathrm{syn}})$, $\gamma = \gamma(P^{\mathrm{syn}})$ and $r = \min\{\rho, \gamma\kappa_\rho^{1/2}/(\beta\bar{y})\}$.*

(a) Fixed integrand. *For every f, such as payment, completions, uplift, an award indicator, or stability in a band with a fixed centre,*

$$\mathbb{E}|\mathbb{E}_{P^{\mathrm{mm}}} f - \mathbb{E}_{P^{\mathrm{syn}}} f| \leq \mathrm{osc}(f)\big\{\sqrt{d}/(4\sqrt{\kappa_\rho n}) + 2de^{-2n\rho^2/d}\big\},$$

and for S independent draws from P^{mm},

$$\mathbb{E}|\mathbb{E}_{P^{\mathrm{int}}} f - \mathbb{E}_{P^{\mathrm{mm}}} f| \leq \mathrm{osc}(f)/(2\sqrt{S}).$$

Summed over the finite menu, $\mathbb{E}\varepsilon_{\mathrm{mm}} = O(n^{-1/2})$ and $\mathbb{E}\varepsilon_{\mathrm{int}} = O(S^{-1/2})$ for these entries; exact summation gives $\varepsilon_{\mathrm{int}} = 0$.

(b) Centred band with a margin. *If $\gamma > 0$, then*

$$\mathbb{E}|F(P^{\mathrm{mm}}) - F(P^{\mathrm{syn}})| \leq \bar{y}\big\{\sqrt{d}/(2\sqrt{\kappa_\rho n}) + 2de^{-2nr^2/d}\big\}.$$

Given P^{mm} with margin $\gamma' > 0$,

$$\mathbb{E}|F(P^{\mathrm{int}}) - F(P^{\mathrm{mm}})| \leq \bar{y}\big\{1/\sqrt{S} + 2e^{-2S\gamma'^2/(\beta\bar{y})^2}\big\}.$$

(c) No margin. *If $\gamma = 0$, $\mathrm{Cov}_{P^{\mathrm{syn}}}(\phi, y) \neq 0$ and $\mathrm{Cov}_P(\phi)$ is nonsingular, then, as $n \to \infty$,*

$$\mathbb{E}|F(P^{\mathrm{mm}}) - F(P^{\mathrm{syn}})| \to \tfrac{1}{2}M^\star(w_- + w_+) > 0,$$

where w_- and w_+ are the P^{syn}-masses of the trajectories that pay exactly $\alpha M^\star$ and $\beta M^\star$.

For an offer whose tilt is interpolated between piloted offers, (a) and (b) hold with the constants of Lemma 9(c).

Under (b) the rate is $n^{-1/2}$, but it sets in only once $n \gtrsim d\beta^2\bar{y}^2/(\kappa_\rho\gamma^2)$, so a small margin delays it. Neither (b) nor (c) affects the upper bound of Proposition 1(i), which uses the measured table errors, whatever their size. Centring the band at a quantity fixed before estimation, such as the requested spend share or the mean issued with the plan, turns the score into a fixed-integrand entry covered by (a).

Lemma 9. Moment-matching term in expectation. *Let $g_f(m) = \mathbb{E}_{Q_{\lambda(m)}} f$. (a) For $m, m' \in \bar{B}(m^\star, \rho)$,*

$$|g_f(m') - g_f(m)| \leq \tfrac{1}{2}\mathrm{osc}(f)\, \kappa_\rho^{-1/2} \|m' - m\|_2.$$

(b) $\mathbb{E}\|\widehat{m} - m^\star\|_2 \leq \sqrt{d/(4n)}$, and $\Pr(\|\widehat{m} - m^\star\|_2 > s) \leq 2de^{-2ns^2/d}$ for $s > 0$. (c) Let an offer's tilt be $\lambda_t = (1-t)\lambda_j + t\lambda_k$ for piloted neighbours j, k, where $\lambda = 0$ is exact at the no-offer plan, and let both pilot means lie in their balls, with κ_ρ the smaller of the two constants. Then

$$|\mathbb{E}_{Q_{\widehat{\lambda}_t}} f - \mathbb{E}_{Q_{\lambda_t^\star}} f| \leq \tfrac{\sqrt{d}}{4}\mathrm{osc}(f)\, \kappa_\rho^{-1} \max_{i \in \{j,k\}} \|\widehat{m}_i - m_i^\star\|_2.$$

Proof. (a) By the chain rule, $\nabla g_f(m) = \Sigma^{-1}c$ with $\Sigma = \Sigma_{\lambda(m)}$ and $c = \mathrm{Cov}_{Q_{\lambda(m)}}(\phi, f)$. Since $c^\top \Sigma^{-1} c$ is the variance of the best linear predictor of f from ϕ,

$$\begin{aligned}\|\Sigma^{-1}c\|_2^2 &\le \kappa_\rho^{-1} c^\top \Sigma^{-1} c \\ &\le \kappa_\rho^{-1}\mathrm{Var}(f) \\ &\le \kappa_\rho^{-1}\mathrm{osc}(f)^2/4.\end{aligned}$$

Integrate along the segment from m to m', which stays in the ball.

(b) Since $\phi \in [0,1]^d$,

$$\begin{aligned}\mathbb{E}\|\widehat{m} - m^\star\|_2^2 &= \mathrm{tr}\,\mathrm{Cov}_P(\phi)/n \\ &\le d/(4n).\end{aligned}$$

For the tail apply Hoeffding's inequality (Hoeffding, 1963) to each coordinate at level $s/\sqrt{d}$ and take a union bound.

(c) *Step 1: covariance bound.* For $\phi \in [0,1]^d$,

$$\Sigma \preceq (\mathrm{tr}\,\Sigma) I \preceq \tfrac{d}{4} I.$$

Along $\lambda_t^\star + \tau u$ with $u = \widehat{\lambda}_t - \lambda_t^\star$, the derivative of the expectation is $\mathrm{Cov}(f, u^\top\phi)$, and

$$\begin{aligned}|\mathrm{Cov}(f, u^\top\phi)| &\le \tfrac{1}{2}\mathrm{osc}(f)(u^\top \Sigma u)^{1/2} \\ &\le \tfrac{1}{2}\mathrm{osc}(f)\tfrac{\sqrt{d}}{2}\|u\|.\end{aligned}$$

Step 2: integrate along the segment. By convexity,

$$\|u\| \le \max_i \|\widehat{\lambda}_i - \lambda_i^\star\|.$$

Integrating $d\lambda/dm = \Sigma^{-1}$ along the segment from $m_i^\star$ to $\widehat{m}_i$ gives

$$\|\widehat{\lambda}_i - \lambda_i^\star\| \le \kappa_\rho^{-1}\|\widehat{m}_i - m_i^\star\|.$$

□

Lemma 10. Stability of the centred band. *Let q, q' be laws on $\mathcal{Z}$. (i) If $\beta|M(q') - M(q)| < \gamma(q)$, then $S(q') = S(q)$ and*

$$|F(q') - F(q)| \le |M(q') - M(q)| + M(q)\,|q'(S(q)) - q(S(q))|.$$

(ii) Let q have full support and $M(q) > 0$, let Z_- and Z_+ be the trajectories paying exactly $\alpha M(q)$ and $\beta M(q)$, and let $w_\pm = q(Z_\pm)$. If $q_k \to q$ with $M(q_k) > M(q)$ for all k, then

$$F(q_k) \to M(q)\{q(S(q)) - w_-\};$$

if $M(q_k) < M(q)$ for all k, then $F(q_k) \to M(q)\{q(S(q)) - w_+\}$.

Proof. (i) Write $M = M(q)$, $M' = M(q')$, $\gamma = \gamma(q)$ and $\Delta = |M' - M| < \gamma/\beta$. A trajectory in $S(q)$ has $\alpha M + \gamma \le y \le \beta M - \gamma$, so

$$\begin{aligned}\alpha M' \le \alpha M + \alpha\Delta < \alpha M + \gamma \le y,\\ y \le \beta M - \gamma < \beta M - \beta\Delta \le \beta M'.\end{aligned}$$

A trajectory outside $S(q)$ has either

$$y \le \alpha M - \gamma < \alpha M'$$

or

$$y \ge \beta M + \gamma > \beta M'.$$

Hence $S(q') = S(q) =: S$ and

$$F(q') - F(q) = (M' - M)\,q'(S) + M\{q'(S) - q(S)\}.$$

(ii) A trajectory with $y \notin \{\alpha M, \beta M\}$ keeps its membership once $\beta|M(q_k) - M|$ is below its distance to the endpoints, as in (i). If $M(q_k) > M$, then $\alpha M(q_k) > y$ on Z_- and $\beta M(q_k) > y$ on Z_+, so eventually $S(q_k) = S(q) \setminus Z_-$ and

$$F(q_k) = M(q_k)\,q_k(S(q) \setminus Z_-) \to M\{q(S(q)) - w_-\}.$$

The other case is symmetric. □

Proof of Corollary 1. (a) On $\{\|\widehat{m} - m^\star\| \le \rho\}$, Lemma 9(a,b) bounds the expected difference by $\frac{1}{2}\mathrm{osc}(f)\kappa_\rho^{-1/2}\sqrt{d/(4n)}$. Off this event the difference is at most $\mathrm{osc}(f)$, and the event has probability at most $2de^{-2n\rho^2/d}$. Given P^{mm}, $\mathbb{E}_{P^{\mathrm{int}}} f$ averages S independent values with variance at most $\mathrm{osc}(f)^2/4$. The interpolated case uses Lemma 9(c) in the same way.

(b) *Moment term.* On $G = \{\|\widehat{m} - m^\star\| \le r\}$, Lemma 9(a) with $f = y$ gives $\beta|M(P^{\mathrm{mm}}) - M^\star| \le \gamma/2 < \gamma$. Lemma 10(i), with Lemma 9(a) for $f = y$ and $f = \mathbf{1}_{S(P^{\mathrm{syn}})}$, then gives

$$\begin{aligned}|F(P^{\mathrm{mm}}) - F(P^{\mathrm{syn}})| &\le \tfrac{1}{2}(\bar{y} + M^\star)\kappa_\rho^{-1/2}\|\widehat{m} - m^\star\| \\ &\le \bar{y}\kappa_\rho^{-1/2}\|\widehat{m} - m^\star\|.\end{aligned}$$

Off G the difference is at most $\bar{y}$, since $0 \le F \le \bar{y}$.

Integration term. Hoeffding's inequality gives

$$\Pr\big(\beta|M(P^{\mathrm{int}}) - M(P^{\mathrm{mm}})| \ge \gamma'\big) \le 2e^{-2S\gamma'^2/(\beta\bar{y})^2};$$

otherwise Lemma 10(i) applies, with $\mathbb{E}|\Delta M| \le \bar{y}/(2\sqrt{S})$ and $\mathbb{E}|\Delta q(S)| \le 1/(2\sqrt{S})$.

(c) *Step 1: asymptotic normality.* Almost surely $\widehat{m} \to m^\star$, and $\sqrt{n}(\widehat{m} - m^\star)$ is asymptotically normal with covariance $\mathrm{Cov}_P(\phi)$.

Step 2: delta method. The gradient of M with respect to the moments at $m^\star$ is $v = \Sigma^{-1}\mathrm{Cov}_{P^{\mathrm{syn}}}(\phi, y) \neq 0$, so $\sqrt{n}\{M(P^{\mathrm{mm}}) - M^\star\}$ is asymptotically normal with variance $v^\top \mathrm{Cov}_P(\phi)v > 0$, and each sign has limiting probability $\frac{1}{2}$.

Step 3: the limit. By Lemma 10(ii) and continuity of $m \mapsto Q_{\lambda(m)}$, with $\Delta M = M(P^{\mathrm{mm}}) - M^\star$,

$$|F(P^{\mathrm{mm}}) - F(P^{\mathrm{syn}})| - M^\star\{w_-\mathbf{1}(\Delta M > 0) + w_+\mathbf{1}(\Delta M < 0)\} \to 0$$

in probability. The difference is bounded by $\bar{y}$, so its expectation converges to $\frac{1}{2}M^\star(w_- + w_+)$. □

Two examples. Both fall under (c).

First example. If $Y \in \{1, \frac{3}{2}\}$ with equal probabilities, then $M^\star = \frac{5}{4}$, the band is $[1, \frac{3}{2}]$, and both payment values sit on its endpoints with $w_- = w_+ = \frac{1}{2}$: shifting mass $t \neq 0$ between them gives

$$F = (\tfrac{5}{4} + \tfrac{t}{2})(\tfrac{1}{2} + |t|) \to \tfrac{5}{8},$$

whereas $F = \frac{5}{4}$ at $t = 0$.

Second example. If a unit prize is won with probability $p^\star = \frac{5}{6}$, the prize sits on the upper endpoint $\beta M^\star = 1$ with $w_+ = \frac{5}{6}$, and the limit is $\frac{1}{2} \cdot \frac{5}{6} \cdot \frac{5}{6} = \frac{25}{72}$. This example runs through Eq. (5) with a nonsingular covariance: four trajectories $(C, D) \in \{(1,1), (2,1), (3,2), (4,2)\}$, features $(C/4, D/2)$, a uniform base, a prize for $C \ge 3$, and truth $Q_{\lambda^\star}$ with $\lambda^\star = (0, 2\log 5)$.

K.4 What the lower bounds cover

Proposition 1(ii) is a statement about tables. Its instances perturb the entries of a fixed menu directly, and the truths consistent with the final tables range over all tables within the stage errors, not over laws produced by an I-projection, a finite pilot and finite sampling. The next proposition asks what the statistical chain itself attains.

Proposition 3. Lower bounds attained by the statistical chain. *Take one layer with two plans x, y that pay nothing, so that every portfolio is feasible and $\omega \equiv 0$, and a value integrand h.*

(a) Synthetic stage, constant attained. *Let p_0 and ϕ be as in Appendix K.3, and let $h : \mathcal{Z} \to [0,1]$ not be affine in ϕ. There is $\bar{\varepsilon} > 0$ such that for every $\varepsilon \in (0, \bar{\varepsilon}]$ there are offer laws P_x, P_y with the same I-projection $Q^\star$ and*

$$\mathbb{E}_{P_x} h - \mathbb{E}_{Q^\star} h = \varepsilon = \mathbb{E}_{Q^\star} h - \mathbb{E}_{P_y} h.$$

With population moments, exact summation and exact enumeration, $\varepsilon^v_{\mathrm{syn}} = \varepsilon$ and the other terms vanish, and every deterministic planner that uses the laws only through their ϕ-moments has regret 2ε under one of the two labellings. Thus Eq. (8) *holds with equality for the actual I-projection.*

(b) Moment-matching stage, order attained. *Let $\mathcal{Z} = \{0, 1\}$, $\phi(z) = h(z) = z$ and any full-support base, so that the tilt family contains every nondegenerate Bernoulli law and P^{mm} is the pilot frequency. Let P_x, P_y be Bernoulli$(\frac{1}{2} \pm \delta)$ with $\delta = (3/(128n))^{1/2}$, and let each plan receive n independent pilot trajectories. Every rule that picks a plan from the pilot data has expected regret at least*

$$\delta/2 = (3/(512n))^{1/2}$$

under one of the two labellings.

(c) Integration stage, order attained by sampling only. *The construction of (b), with S independent draws from P^{mm} in place of the pilot, gives the S-draw integrator expected regret at least $(3/(512S))^{1/2}$ on some instance. Exact summation over a finite $\mathcal{Z}$ has no integration error by Lemma 7.*

Proof. (a) *Step 1: a direction invisible to the moments.* Since $h \notin \operatorname{span}\{1, \phi_1, \dots, \phi_d\}$ in $\mathbb{R}^{\mathcal{Z}}$, its residual w after orthogonal projection onto that span is nonzero, with

$$\sum_z w(z) = 0,$$
$$\sum_z w(z)\phi(z) = 0,$$
$$\langle w, h\rangle = \|w\|^2;$$

rescale w so that $\langle w, h\rangle = 1$.

Step 2: the two laws. Let $Q^\star = Q_\lambda$ for any λ and $\bar{\varepsilon} = \min_{w(z)\neq 0} Q^\star(z)/|w(z)|$. Then

$$P_x = Q^\star + \varepsilon w,$$
$$P_y = Q^\star - \varepsilon w$$

are laws with moments $\mathbb{E}_{Q^\star}\phi$. By the identity of Lemma 4, for every q with these moments,

$$\mathrm{KL}(q\|p_0) = \mathrm{KL}(q\|Q^\star) + \mathrm{KL}(Q^\star\|p_0),$$

so $Q^\star$ is the I-projection of both, and their values are $\mathbb{E}_{Q^\star} h \pm \varepsilon$.

Step 3: the regret. All later tables of x and y coincide, so a deterministic planner's choice depends only on the labels, and the labelling under which it picks the plan with law P_y has regret 2ε. A randomized planner has regret at least ε averaged over the two labellings.

(b) *Step 1: two-point reduction.* Let $\mathsf{P}_1, \mathsf{P}_2$ be the laws of the $2n$ pilot outcomes under the two labellings. A rule ψ has regrets $2\delta\,\mathsf{P}_1(\psi = y)$ and $2\delta\,\mathsf{P}_2(\psi = x)$, whose sum is at least

$$2\delta\{1 - \mathrm{TV}(\mathsf{P}_1, \mathsf{P}_2)\}.$$

Step 2: bounding the distance. For $\delta \le \frac{1}{4}$,

$$\mathrm{KL}(\mathsf{P}_1\|\mathsf{P}_2) = 4n\delta \log\frac{1+2\delta}{1-2\delta} \le \frac{64}{3} n\delta^2,$$

so Pinsker's inequality gives

$$\mathrm{TV} \le \Big(\frac{32}{3} n\delta^2\Big)^{1/2} = \frac{1}{2}.$$

(c) This is (b) with the S draws as the data. □

Part (a) shows that the I-projection itself attains the constant 2 of the synthetic term, and that this term is the price of compressing the pilot to d moments: P_x and P_y have different trajectory histograms, which a trial over every offer would observe. Parts (b) and (c) show that the $n^{-1/2}$ and $S^{-1/2}$ orders of Corollary 1 cannot be improved in general: the first by no use of the same pilot, the second only by giving up sampling.

Scope of the decomposition. Four kinds of loss are distinct, and only some are bounded. (1) Proposition 1(i) is a perturbation bound: it compares the planner with the best portfolio over the *same layer menus*. The value lost by restricting individuals to layer menus,

$$\Delta_{\mathrm{menu}} = V^{\star}_{\mathrm{ind}} - V^{\star}(0),$$

with $V^{\star}_{\mathrm{ind}}$ the best true value over individual plans, is not bounded here, and neither are the ranking and roster choices of M1–M2; they are measured only against the layerings that were tested. (2) The loss from tightening the budget, $\omega(2E_c)$, is bounded only under a margin or growth condition: on a finite menu $V^{\star}$ is a step function, so $\omega(2E_c) = 0$ when an optimal portfolio's true spend lies in $\mathcal{B}_{2E_c}$, and $\omega(2E_c) \le 2\kappa E_c$ when $\omega(\delta) \le \kappa\delta$. The tightening itself needs E_c, which is unknown when a plan is issued unless a certificate supplies it, and Algorithm 1 uses the untightened band. (3) Part (ii) of the proposition holds at the level of tables; Proposition 3 is the statement for the chain of laws. (4) Part (iii) and Proposition 3 assert that certain instances exist. They do not say that on a given instance an accurate stage cannot partly offset another: errors can cancel, and the direct error in Table 10 is below the sum of the stage errors. Nor do they say that realized regret grows whenever one stage term grows. That a stage is needed in practice rests on the ablations and holds relative to the replacements tested there.

L REGISTERED TEST FAMILIES

Table 29: Pre-specified decision rules and their outcomes; the rules were fixed before the runs, and except for the integrated study also the analysis scripts. Rider held-out validation, Appendix A.1: paired change, held-out minus reference, of plan-cost error and selection regret at budget fraction 1, 95% intervals over origins; the rule requires upper bounds ≤ 0.045 and ≤ 0.01. Synthetic trajectories and a short pilot, Section 5: the pre-specified primary arm is the framework with sampled integration, $S = 2{,}048$; differences of penalized relative regret at 10% scoring on the 18 response laws, 95% intervals over laws. Integrated study, Appendix C: the registered confirmatory family on 36 response laws; H1 is the framework minus the same-calendar trial in relative regret with the band within 20% of the budget, H2 the geometric-mean speed-up of answering all 60 requests of an origin by reuse, one-off preparation charged, over re-running the whole pipeline per request; 95% intervals over laws. Its analysis script was completed after the run; H2 is computed on the registered 36 response laws. New generator, Appendix D: the same family on its 36 response laws, with the rules and the analysis script fixed before the run. Stronger controls, behaviour change, joint fit, population reduction, departures from the tilt family, finite riders and finite data, studies C–I of Appendix E: registered families of seven, five, five, three, eighteen, four, sixteen and fourteen one-sided tests, the last two for study G continued and for study I, Holm step-down at level 0.025 within each family; estimates with 95% law-bootstrap intervals, relative regret within 20% of the budget, and for study F the retained share of the type-informed continuous-population optimum. Study E: joint fit as specified by the registered amendment in Appendix E.

Held-out condition	Change in plan-cost error	Change in regret	Met
Unseen period, fit once, 16 origins	$+0.051$ [$+0.002$, $+0.098$]	$+0.000$ [$+0.000$, $+0.000$]	no
Unseen cities, four folds, 45 origins	$+0.003$ [-0.018, $+0.020$]	$+0.000$ [$+0.000$, $+0.000$]	yes
Pilot rule	**1-week pilot**	**2-week pilot**	**4-week pilot**
Near-lossless vs. 18-week trial, upper bound ≤ 0.05	$+0.156$ [$+0.114$, $+0.201$] (no)	$+0.093$ [$+0.061$, $+0.126$] (no)	$+0.059$ [$+0.039$, $+0.077$] (no)
Beats same-calendar trial, upper bound < 0	-0.256 [-0.290, -0.222] (yes)	-0.160 [-0.186, -0.134] (yes)	-0.076 [-0.109, -0.043] (yes)
Without history is worse, lower bound > 0	$+0.371$ [$+0.266$, $+0.488$] (yes)	$+0.388$ [$+0.261$, $+0.525$] (yes)	$+0.380$ [$+0.258$, $+0.514$] (yes)
Without moment matching is worse, lower bound > 0	$+0.369$ [$+0.235$, $+0.499$] (yes)	$+0.433$ [$+0.302$, $+0.561$] (yes)	$+0.467$ [$+0.359$, $+0.574$] (yes)
Without joint integration is worse, lower bound > 0	$+0.296$ [$+0.226$, $+0.368$] (yes)	$+0.376$ [$+0.299$, $+0.458$] (yes)	$+0.428$ [$+0.374$, $+0.487$] (yes)
Without budget allocation is worse, lower bound > 0	$+0.581$ [$+0.516$, $+0.652$] (yes)	$+0.643$ [$+0.568$, $+0.719$] (yes)	$+0.676$ [$+0.612$, $+0.737$] (yes)
Integrated-study rule	**1-week pilot**	**2-week pilot**	**4-week pilot**
H1: beats same-calendar trial within 20%, upper bound < 0	-0.171 [-0.201, -0.140] (yes)	-0.124 [-0.160, -0.089] (yes)	-0.069 [-0.104, -0.035] (yes)
H2: 60 requests faster with preparation, lower bound > 1, same answers	$14.1\times$ [13.8, 14.5] (yes)	$14.2\times$ [13.8, 14.5] (yes)	$14.3\times$ [14.0, 14.7] (yes)
New-generator rule	**1-week pilot**	**2-week pilot**	**4-week pilot**
H1: beats same-calendar trial within 20%, upper bound < 0	-0.289 [-0.311, -0.267] (yes)	-0.176 [-0.192, -0.160] (yes)	-0.123 [-0.140, -0.108] (yes)
H2: 60 requests faster with preparation, lower bound > 1, same answers	$2.70\times$ [2.67, 2.73] (yes)	$2.72\times$ [2.69, 2.75] (yes)	$2.78\times$ [2.76, 2.81] (yes)

Table 30: Pre-specified decision rules and their outcomes, continued from Table 29: studies C–H, registered families with Holm step-down at level 0.025 within each family; estimates with 95% law-bootstrap intervals. Study G: the framework minus the joint fit or the same-calendar trial at one week; strengths 1–3 are $\pi = 1/2, 1/4, 1/8$ for the tail, $\delta = 0.10, 0.25, 0.50$ for the support gap and $\kappa = 0.25, 0.5, 1$ for the curvature. Study H: 20 riders per group. H6.5 gives tied values their average rank, as registered; ranking them in input order gives the same value to three decimals.

Stronger-control rule	1-week pilot	2-week pilot	4-week pilot
H6.1: tilted trial minus trial, < 0	-0.075 [-0.078, -0.072] (yes)		
H6.2: minus dose interpolation, < 0	-0.022 [-0.024, -0.021] (yes)		
H6.3: minus normal, exact truncated moments, < 0	-0.053 [-0.055, -0.051] (yes)		
H6.4: calendar shocks, minus trial, < 0	-0.267 [-0.274, -0.259] (yes)	-0.214 [-0.222, -0.206] (yes)	
H6.5: ESS deficit vs. cost error, Spearman, ties at average rank, > 0	$+0.415$ [$+0.400$, $+0.431$] (yes)		
H6.6: support 1,024 minus 16,384 weeks, > 0	$+0.006$ [$+0.006$, $+0.006$] (yes)		
Behaviour-change rule	**1-week pilot**	**2-week pilot**	**4-week pilot**
H7.1–H7.3: request-aware minus direct reuse, < 0	-0.195 [-0.202, -0.189] (yes)	-0.255 [-0.261, -0.249] (yes)	-0.300 [-0.306, -0.293] (yes)
H7.4: same, selected subpopulations, < 0	-0.347 [-0.356, -0.338] (yes)		
H7.5: request-aware minus trial, 288 laws, < 0	-0.129 [-0.136, -0.123] (yes)		
Joint-fit rule, study E	**1-week pilot**	**2-week pilot**	**4-week pilot**
H9.1: framework minus joint fit of the trial, < 0	-0.055 [-0.059, -0.051] (yes)	-0.052 [-0.056, -0.048] (yes)	-0.060 [-0.066, -0.053] (yes)
H9.2: joint fit minus per-offer tilt of the trial, < 0	-0.152 [-0.158, -0.146] (yes)		
H9.3: cost-table error, framework minus joint fit, < 0	-0.014 [-0.015, -0.014] (yes)		
Reduction rule, study F	**retained share of the type-informed optimum, difference**		
H8.1: five layers minus group-level plan, > 0	$+0.120$ [$+0.116$, $+0.125$] (yes)		
H8.2: five quantile layers minus five equal-width bins, > 0	-0.002 [-0.002, -0.001] (no)		
H8.3: five layers minus two layers, > 0	$+0.046$ [$+0.044$, $+0.047$] (yes)		
Departure rule, study G	**strength 1**	**strength 2**	**strength 3**
H10.T: tail, minus joint fit, < 0	-0.055 [-0.058, -0.052] (yes)	-0.054 [-0.057, -0.050] (yes)	-0.065 [-0.069, -0.061] (yes)
H10.T: tail, minus trial, < 0	-0.286 [-0.293, -0.279] (yes)	-0.293 [-0.299, -0.286] (yes)	-0.302 [-0.307, -0.297] (yes)
H10.S: support gap, minus joint fit, < 0	-0.046 [-0.049, -0.043] (yes)	-0.016 [-0.021, -0.011] (yes)	-0.078 [-0.081, -0.075] (yes)
H10.S: support gap, minus trial, < 0	-0.244 [-0.250, -0.238] (yes)	-0.148 [-0.157, -0.138] (yes)	$+0.329$ [$+0.312$, $+0.345$] (no)
H10.K: curvature, minus joint fit, < 0	-0.024 [-0.030, -0.018] (yes)	$+0.025$ [$+0.017$, $+0.034$] (no)	$+0.071$ [$+0.065$, $+0.077$] (no)
H10.K: curvature, minus trial, < 0	-0.138 [-0.148, -0.127] (yes)	$+0.053$ [$+0.041$, $+0.066$] (no)	$+0.325$ [$+0.315$, $+0.336$] (no)
Finite-rider rule, study H	**retained share, difference**		
H8b.1: five layers minus group-level plan, type-informed integer optimum, > 0	$+0.126$ [$+0.118$, $+0.135$] (yes)		
H8b.2: five layers, same-information minus type-informed integer optimum, > 0	$+0.024$ [$+0.022$, $+0.025$] (yes)		
H8b.3: the same in the continuum, > 0	$+0.035$ [$+0.033$, $+0.037$] (yes)		
H8b.4: five layers, same-information integer optimum minus 0.95, > 0	$+0.034$ [$+0.032$, $+0.035$] (yes)		

Table 31: Pre-specified decision rules and their outcomes, continued from Table 30, with Holm step-down at level 0.025 within each family; estimates with 95% law-bootstrap intervals. Study G, continued: the framework minus the joint fit at one week, under the uniform rule for unreachable targets with every fit converged, H11.P, and with 2,000 and 5,000 steps under the output rule of study G, H11.L; strengths as in Table 30. Study I: retained share with L layers minus one layer at the same pilot and history, within 20% of the budget.

Converged-fit rule, study G	strength 1	strength 2	strength 3
H11.P.M0: uniform rule, M0, < 0	-0.056 [-0.060, -0.052] (yes)		
H11.P.T: uniform rule, tail, < 0	-0.055 [-0.058, -0.052] (yes)	-0.054 [-0.057, -0.050] (yes)	-0.065 [-0.069, -0.061] (yes)
H11.P.S: uniform rule, support gap, < 0	-0.046 [-0.049, -0.043] (yes)	-0.015 [-0.019, -0.010] (yes)	$+0.001$ [-0.001, $+0.003$] (no)
H11.P.K: uniform rule, curvature, < 0	-0.024 [-0.030, -0.018] (yes)		
H11.L.M0: 2,000 and 5,000 steps, M0, < 0	-0.056 [-0.060, -0.052] (yes)		
H11.L.T: 2,000 and 5,000 steps, tail, < 0	-0.055 [-0.058, -0.052] (yes)	-0.054 [-0.057, -0.050] (yes)	-0.065 [-0.069, -0.061] (yes)
H11.L.S: 2,000 and 5,000 steps, support gap, < 0	-0.046 [-0.049, -0.043] (yes)	-0.007 [-0.013, -0.002] (yes)	-0.098 [-0.102, -0.093] (yes)
H11.L.K: 2,000 and 5,000 steps, curvature, < 0	-0.024 [-0.030, -0.018] (yes)		
Layer rule, study I	**rider-weeks per group n; retained share, L layers minus one layer**		
H12.L3: three layers, $n = 135, 270, 540, > 0$	$+0.060$ [$+0.057$, $+0.063$] (yes)	$+0.081$ [$+0.078$, $+0.084$] (yes)	$+0.086$ [$+0.083$, $+0.089$] (yes)
H12.L3: $n = 1{,}080$, 2,160, 4,320, > 0	$+0.086$ [$+0.083$, $+0.090$] (yes)	$+0.079$ [$+0.076$, $+0.083$] (yes)	$+0.073$ [$+0.069$, $+0.077$] (yes)
H12.L3: $n = 8{,}640, > 0$	$+0.066$ [$+0.062$, $+0.070$] (yes)		
H12.L5: five layers, $n = 135, 270, 540, > 0$	$+0.026$ [$+0.023$, $+0.030$] (yes)	$+0.077$ [$+0.074$, $+0.080$] (yes)	$+0.095$ [$+0.091$, $+0.098$] (yes)
H12.L5: $n = 1{,}080$, 2,160, 4,320, > 0	$+0.102$ [$+0.098$, $+0.106$] (yes)	$+0.099$ [$+0.095$, $+0.103$] (yes)	$+0.092$ [$+0.087$, $+0.096$] (yes)
H12.L5: $n = 8{,}640, > 0$	$+0.079$ [$+0.075$, $+0.084$] (yes)		

Table 32: Pre-specified decision rules and their outcomes, continued from Table 31, with Holm step-down at level 0.025 within each family; estimates with 95% law-bootstrap intervals, relative regret within 20% of the budget. Study J of Appendix F: the framework minus each learned pipeline. Study K: the operating policy OP minus the framework at M0 with non-inferiority margin 0.02, minus the same-calendar trial, minus its fallback with margin 0.02 and minus the framework under curvature; one-week pilot, strengths as in Table 30.

Learned-pipeline rule, study J	1-week pilot	2-week pilot	4-week pilot
H14.LDR: minus learned density ratio, < 0	-0.356 [-0.364, -0.347] (yes)	-0.385 [-0.394, -0.376] (yes)	-0.344 [-0.354, -0.335] (yes)
H14.RM: minus response-model pipeline, < 0	-0.020 [-0.022, -0.018] (yes)	-0.017 [-0.018, -0.015] (yes)	-0.011 [-0.012, -0.009] (yes)

Operating-policy rule, study K	strength 1	strength 2	strength 3
H15.RET.M0: minus framework, M0, < 0.02	$+0.052$ [$+0.048$, $+0.055$] (no)		
H15.SC.M0: minus trial, M0, < 0	-0.228 [-0.234, -0.222] (yes)		
H15.SC.T: minus trial, tail, < 0	-0.227 [-0.234, -0.221] (yes)	-0.220 [-0.226, -0.214] (yes)	-0.214 [-0.220, -0.208] (yes)
H15.SC.S: minus trial, support gap, < 0	-0.148 [-0.153, -0.143] (yes)	-0.107 [-0.112, -0.103] (yes)	-0.097 [-0.101, -0.092] (yes)
H15.SC.K: minus trial, curvature, < 0	-0.121 [-0.129, -0.113] (yes)	-0.015 [-0.022, -0.008] (yes)	$+0.048$ [$+0.044$, $+0.052$] (no)
H15.SAFE.M0: minus fallback, M0, < 0.02	-0.134 [-0.137, -0.130] (yes)		
H15.SAFE.T: minus fallback, tail, < 0.02	-0.130 [-0.134, -0.126] (yes)	-0.118 [-0.122, -0.114] (yes)	-0.108 [-0.112, -0.103] (yes)
H15.SAFE.S: minus fallback, support gap, < 0.02	-0.052 [-0.054, -0.049] (yes)	-0.011 [-0.012, -0.009] (yes)	$+0.000$ [$+0.000$, $+0.000$] (yes)
H15.SAFE.K: minus fallback, curvature, < 0.02	-0.040 [-0.045, -0.035] (yes)	$+0.053$ [$+0.048$, $+0.057$] (no)	$+0.098$ [$+0.094$, $+0.101$] (no)
H15.REP.K: minus framework, curvature, < 0		-0.069 [-0.075, -0.062] (yes)	-0.278 [-0.288, -0.268] (yes)