SCSV: SUBSTITUTION-CORRECTED SUPPLY VALUATION FOR VALUE-GUIDED INTERVENTION IN RECOMMENDER ECOSYSTEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Recommender platforms continuously invest attention in new items and creators, yet under finite attention the engagement an item attracts may be reallocated from substitutes rather than created. Engagement-based estimators overlook this substitution, while randomized shielding is difficult to scale to item-level valuation. We propose Substitution-Corrected Supply Value (SCSV), which values an item by a shielding estimand, the ecosystem utility lost when it is removed and attention reallocates to substitutes. SCSV settles this value from production logs, using lagged differencing against exposure leakage and double machine learning against observed confounding. Its settlement is sign-gated and replaceability-discounted, and caps each attribution unit's credit by its realized utility change. The intervention layer distills settled credits into uncertainty-aware priors, ranks candidates by risk-adjusted value density, and allocates exposure only when gains exceed displaced value, with conditional bounds on selection optimality and allocation error. We construct the first industrial-scale benchmark pairing production logs with randomized supply-shielding ground truth. Set-level estimates agree with these shielding losses in sign and ordering, within 0.3 pp on user-initiated content entries (open reasons) and with dose-dependent gaps on watch time. SCSV improves selection efficiency over incumbent metrics by 226%, 90.3%, and 46.9% in cold-start, production, and live-streaming scenarios. In online experiments with at least 20 million daily active users, SCSV improves watch time, open reasons, and long-term retention, while raising open-reason intervention ROI to 3.6 times the baseline. These results support substitution-corrected valuation for directing supply investment toward utility that is created rather than reallocated.

1 INTRODUCTION

Recommender platforms continuously invest finite user attention in new supply across an evolving inventory, through cold-start ramp-up and positive supply intervention, and must decide which supply deserves it. The engagement an item attracts is the natural answer. Because supply is substitutable, however, that engagement may be reallocated from substitutes rather than created. The item-attributed engagement that ranking objectives and click-through-rate (CTR) models optimize (Wang et al., 2021; Mao et al., 2023) therefore does not measure the item's marginal contribution to the ecosystem (Fig. 1). We call the utility that substitutes lose when an item gains exposure cannibalization, and the utility they recover when it loses exposure compensation. Substitution is interference among items. An item's outcome depends on the exposure of others, violating the Stable Unit Treatment Value Assumption (SUTVA) with items as units (Imbens & Rubin, 2015; Yao et al., 2023). An item's contribution therefore depends on the supply that remains and on how attention is reallocated, and it must be defined as an ecosystem-level counterfactual rather than read off the item's own feedback (Mladenov et al., 2020). That counterfactual is the utility the ecosystem loses when the item is removed and the serving policy reallocates its attention to substitutes.

Three obstacles separate this removal value from a deployable signal. ***Estimation.*** In production logs, ranking decisions shape what users see and responses feed back into future exposure, so attribution faces exposure leakage and high-dimensional confounding. Even a deconfounded attribution still credits an item with utility that substitutes would have recovered, because logs never realize the

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.

post-removal reallocation for an individual item. Credits summed within a user also need not respect the utility change that user actually realized. Item-attributed feedback must therefore be separated from allocation-induced confounding and from demand reallocation before it can be read as marginal contribution. Randomized shielding provides a direct experimental reference for removal effects (Kohavi et al., 2020). At the item level, however, shielding long-tail supply is underpowered under audience heterogeneity and rapid demand reallocation. Arms that share supply interfere (Blake & Coey, 2014; Johari et al., 2022), each round may carry an experience cost, and an item that has not yet been served has no removal effect to measure. Platforms therefore randomize over item sets or clusters (Saint-Jacques et al., 2019). Set-level randomization makes the removal target testable but does not resolve the contribution of each item, and correlational metrics resolve items but ignore substitution. The remaining challenge is to settle from logs item-level credits that discount recoverable utility, stay within realized ecosystem change, and aggregate to the set-level effects that shielding experiments can test. ***Allocation.*** Even a valid item-level value does not by itself produce ecosystem gains. Positive supply intervention must convert an item-level value signal into platform-level gains by allocating finite attention while accounting for exposure costs, the value displaced from substitutes, and the shared attention budget. Boosted items displace one another, so returns diminish, and ramp-up rules driven by item-level metrics are tied to no net-gain objective. Cold-start supply must moreover be selected before any credit is settled, from predictions whose reliability varies across items. ***Evaluation.*** Public resources offer semi-synthetic simulators (Ie et al., 2019) or logged and exposure-intervention datasets for off-policy and exposure-bias evaluation (Saito et al., 2021; Gao et al., 2022a;b). None pairs production logs with randomized supply-shielding outcomes, so the system-level effect of removing items or creators remains unobserved.

To address these challenges, we present Substitution-Corrected Supply Value (SCSV), which values an item by this removal counterfactual. Our contributions are as follows. **Estimand and settlement.** We formalize this value as a shielding estimand and settle from production logs a credit that targets it. Attribution is lagged against exposure leakage and orthogonalized (DML) against confounding. Settlement is sign-gated and replaceability-discounted, and caps the credit of each attribution unit by its realized utility change (Proposition 2.1). **Value-guided intervention.** We distill settled credit into an uncertainty-aware prior for cold-start supply and formulate value-guided intervention under finite attention as density-ordered selection and displacement-aware allocation, with conditional bounds on selection optimality and allocation error. **Benchmark.** We construct the first industrial-scale benchmark pairing production logs with randomized supply-shielding ground truth, with an initial package in this submission and the full release upon acceptance. **Validation and deployment.** On this benchmark, set-level SCSV estimates track randomized shielding losses in sign and ordering, within 0.3 pp on open reasons and with dose-dependent gaps on watch time. Deployed as a prior-based gate and boost for at least 20 million daily active users, SCSV raises selection efficiency over incumbent value metrics in three scenarios and lifts intervention ROI on open reasons to 3.6 times the baseline.

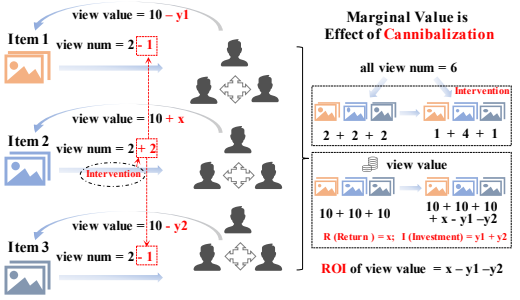


Figure 1: Net utility under exposure reallocation. Boosting Item 2 gains x but displaces y_1 and y_2 from Items 1 and 3, a net gain of $x - y_1 - y_2$, labeled in the figure as the ROI of view value (a difference, not the ratio ROI of Section 4).

Related work. Prior work studies recommender ecosystems as multi-stakeholder systems in which ranking shapes user utility and supply-side incentives (Boutilier et al., 2024; Mladenov et al., 2020), quantifies crowding-out under top-K ranking (Yao et al., 2023), separates incremental consumption from redistribution (Aridor et al., 2022; Zielnicki et al., 2025), and constrains uplift to platform-level consistency (He et al., 2026), but none of it yields an item-level value computable from production logs. Off-policy estimators assume SUTVA (Schnabel et al., 2016; Joachims et al., 2017), interference-aware designs randomize over clusters (Hudgens & Halloran, 2008; Saint-Jacques et al., 2019), and double machine learning handles observed confounding (Chernozhukov et al., 2018), whereas SCSV settles deconfounded attribution into substitution-corrected credit. Exposure-allocation work treats exposure as a constrained resource (Singh & Joachims, 2018; Biega et al., 2018) without targeting

net ecosystem utility, and existing benchmarks (Li et al., 2010; Gao et al., 2022b) pair no production logs with randomized supply-shielding outcomes. Appendix M discusses these lines of work in full.

2 METHODOLOGY

This section develops the full methodology, and Fig. 2 summarizes the SCSV estimation pipeline.

2.1 PROBLEM FORMULATION

Let t index production periods and let $\mathcal{I}_t$ denote the available supply inventory. The platform allocates finite user attention through an exposure vector $v_t \in \mathbb{R}_+^{|\mathcal{I}_t|}$, where $v_{i,t}$ is the exposure intensity assigned to item i. The exposure allocation is constrained by $\sum_{i \in \mathcal{I}_t} v_{i,t} \le C_t$. Let U_t denote an ecosystem-level utility metric, such as total watch time, app duration, total comments, or open reasons, where an open reason is a user-initiated entry into content attributed to a revisit, share, follow-page, or search session. Under finite attention, increasing the exposure of one item can alter the exposure and outcomes of other items. Hence, item-attributed engagement is not, in general, a valid measure of marginal ecosystem contribution. Exposure is allocated by the ranker from item-level performance aggregated across users. The treatment studied in Section 2.2 is therefore a supply-side change in an item's exposure share within this conserved budget, not reinforcement of an individual user's preferences (Appendix A).

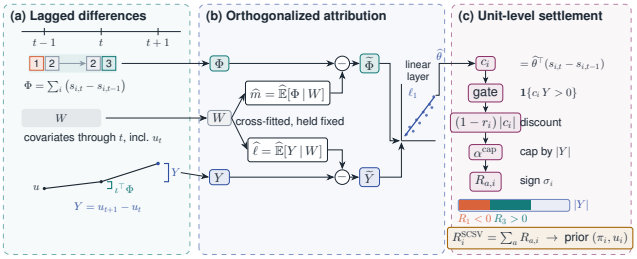


Figure 2: Estimation pipeline with unit and period subscripts suppressed: (a) lagged differences, (b) orthogonalized attribution, whose linear attribution layer gives $\widehat{\theta}$, and (c) unit-level settlement into signed credits within the unit budget $|Y|$.

Shielding estimand. We define substitution-corrected value through a shielding counterfactual. SCSV measures how next-period utility responds to the period-t allocation, the lagged design of Section 2.2, so the outcome window of every quantity below is period $t+1$. The same-period response is outside the estimand. In the settlement logs the period-t utility is the value-weighted behavior of period t (Section 2.2), so it moves with the period-t allocation. Under a shielding policy it is the model-free loss of the first shielded day, which enters the next-period loss as a separate term (Appendix B.1). Let $T_{-i}(v_t)$ denote the exposure allocation induced when item i is shielded and the ranker reallocates the released attention to the remaining supply. The map T_{-i} is determined by the serving policy, so every quantity built on it is policy-dependent. The shielding effect of item i is

$$V_{i,t} = \mathbb{E}\big[U_{t+1}(v_t) - U_{t+1}\big(T_{-i}(v_t)\big) \mid \mathcal{H}_t\big], \tag{1}$$

where $U_{t+1}(v)$ is the ecosystem utility of period $t+1$ when the period-t allocation is v, and $\mathcal{H}_t$ denotes the information available when the period-t allocation is decided. This effect is the causal target. It measures the ecosystem utility lost when item i is removed under the serving policy, net of the compensation provided by substitute supply. It depends on the reallocation, on the supply that remains, and on the outcome window, and it can be negative for an item whose removal frees attention for better substitutes. The gross alternative $U_{t+1}(v_t) - U_{t+1}(v_t - v_{i,t}e_i)$ removes item i without reallocating its exposure. Under additive attribution it equals the utility attributed to i itself and can be read from logs, but it credits i with demand that substitutes would have recovered. The net quantity instead requires the post-substitution allocation $T_{-i}(v_t)$, which observational logs never realize for an arbitrary item and randomized shielding realizes in practice only at the cohort level. Appendix A tabulates the three utilities and gives a worked example.

Three objects. We distinguish the policy-dependent shielding effect from the operational credit used for supply intervention. The shielding effect $V_{i,t}$ is the causal target. The settled credit $R_{i,t}^{\mathrm{SCSV}}$ of Section 2.2 is an operational, signed quantity built from logs by deconfounded attribution, a replaceability discount, and the settlement constraint below. The serving-time prior $\pi_{i,t}$ predicts the settled credit from upload-time features (Appendix D), and Section 2.3 uses its risk-adjusted form $P_{i,t}$. We do not claim that $R_{i,t}^{\mathrm{SCSV}}$ is an item-level consistent estimator of $V_{i,t}$, and we introduce no

such estimator. The agreement between settled credit and shielding effects is instead tested at the level of shielded sets in Section 3.

Credit settlement constraint. Settlement is constrained separately from the estimand. Let a index attribution units, such as users, and let $Y_{a,t+1}$ be the realized next-period utility change of unit a (Section 2.2). Since attributed engagement may include demand reallocated from other items, the credit settled within a unit is not allowed to exceed the change that the unit realized. We therefore require $\sum_i |R_{a,i,t}| \leq |Y_{a,t+1}|$ for every unit, where $R_{a,i,t}$ is the credit that unit a settles on item i and $R^{\mathrm{SCSV}}_{i,t} = \sum_a R_{a,i,t}$. This constraint bounds each unit's total, not the credit of any single item, and distinguishes value settlement from exposure allocation. The settlement budget is an accounting constraint and does not, by itself, identify shielding effects. The quantity $Y_{a,t+1}$ is an observed temporal change, not a causal effect of supply, so zero credit does not mean zero supply value. What settlement preserves is the within-unit ordering of items by discounted attributed mass (Proposition 2.1). Appendix A lists the value objects (Table 8) and works through this case.

The settled credit, through its serving-time prior, feeds a value-guided intervention problem. A support action assigns exposure $h_{i,t} \geq 0$ to a candidate item and draws it from the inventory through fixed displacement weights $\omega^{+}_{ij,t}$, the fraction of the action's exposure taken from item j, so the joint change reallocates existing exposure within the surface cap C_t, while a separate cap C^{+}_t limits the gross support exposure. The platform seeks jointly feasible actions that maximize the expected next-period utility gain, and the relevant gain of a single action is its incremental return net of the utility lost on the supply it displaces. Appendix E states this objective in full (Eqs. (34) and (35)), and Section 2.3 constructs a score-based approximation to it.

2.2 Leakage-resistant Deconfounded Attribution and Credit Settlement

Estimating item value from same-period exposure and utility is vulnerable to exposure-induced leakage. Ranking decisions determine exposure, while user responses feed back into subsequent allocation. To avoid this contemporaneous feedback path, SCSV constructs lagged temporal differences and attributes future utility changes to preceding changes in exposure.

Let a index an attribution unit, such as a user, user-surface pair, or behavior cohort. Let $u_{a,t}$ denote the utility of unit a in period t, computed from the value-weighted behavior signals of Appendix H. Let $b_{a,i,t} \in [0,1]$ denote the exposure share of item i within unit a in period t, the fraction of the unit's attention budget allocated to item i, so that $\sum_i b_{a,i,t} = 1$. The item-level exposure intensity of Section 2.1 is then $v_{i,t} = \sum_a b_{a,i,t} C_{a,t}$, with $C_{a,t}$ the attention volume of unit a. Let $s_{a,i,t}$ denote the vector of value-weighted behavior signals that unit a records on item i in period t (Appendix H), zero when the item is not in the unit's feed. We define

$$Y_{a,t+1} = u_{a,t+1} - u_{a,t}, \qquad X_{a,i,t} = b_{a,i,t} - b_{a,i,t-1}, \tag{2}$$

and restrict attribution to temporally changed items, $\mathcal{C}_{a,t} = \{i : s_{a,i,t} \neq s_{a,i,t-1}\}$, the items on which the unit's behavior vector changed. Unchanged items receive no direct credit. Each sample spans three consecutive periods, with the exposure change measured from $t-1$ to t and the outcome change from t to $t+1$. The utility is the value-weighted total of the behavior vectors, $u_{a,t} = \iota^\top \sum_i s_{a,i,t}$ with ι the fixed production weights (the vector of ones when the signals are stored already weighted), so $u_{a,t} - u_{a,t-1} = \iota^\top \Phi_{a,t}$ holds as an identity, the increment marked $\iota^\top \Phi$ in Fig. 2(a). The same-period utility change is the treatment, and only the next-period change is a response to it.

Let $\phi_{a,i,t} = s_{a,i,t}/b_{a,i,t}$ denote the item's per-share behavior profile in period t, and, for an item that has left the feed, $b_{a,i,t} = 0$, its profile in period $t-1$. Let $W_{a,t}$ denote observed user, traffic, surface, and system-allocation confounders. The treatment term of item i is the change in its behavior vector, $s_{a,i,t} - s_{a,i,t-1}$, and we aggregate sparse treatment variation as $\Phi_{a,t} = \sum_{i \in \mathcal{C}_{a,t}} (s_{a,i,t} - s_{a,i,t-1})$, so an item that enters the feed contributes its period-t vector and an item that leaves the negative of its period-$(t-1)$ vector. For an item present in only one of the two periods the term equals $X_{a,i,t}\phi_{a,i,t}$, its share change at its profile, which is the supply-side treatment of Section 2.1, and $\Phi_{a,t} = \sum_{i \in \mathcal{C}_{a,t}} X_{a,i,t}\phi_{a,i,t}$ when every changed item is of this kind. For an item present in both periods the term also carries the change in its per-share behavior at its previous share (Appendix B). We use the partially linear model $Y_{a,t+1} = \theta^\top \Phi_{a,t} + f(W_{a,t}) + \varepsilon_{a,t}$, where θ parameterizes attribution weights and $f(\cdot)$ absorbs high-dimensional confounding. Only the attribution layer is linear. The behavior vector $s_{a,i,t}$ may include non-linear transforms of raw signals, and $f(\cdot)$ is fit by cross-fitted

models without functional-form restrictions. We keep θ linear so that unit-level credit remains auditable, and the fitted response is linear in $\Phi_{a,t}$ but not in the policy behind it (Appendix B.1).

SCSV estimates θ by DML-style orthogonalization. DML (Chernozhukov et al., 2018) fits nuisance models of how $W_{a,t}$ predicts $\Phi_{a,t}$ and $Y_{a,t+1}$ on held-out folds and estimates θ from the residuals, so that errors in the nuisance fits enter only at second order. Let $\widetilde{\Phi}_{a,t}$ and $\widetilde{Y}_{a,t+1}$ denote cross-fitted residuals after partialing out $W_{a,t}$ from $\Phi_{a,t}$ and $Y_{a,t+1}$, respectively. We solve

$$\widehat{\theta} \in \arg\min_{\theta} \sum_{a,t} \left(\widetilde{Y}_{a,t+1} - \theta^{\top} \widetilde{\Phi}_{a,t} \right)^2 + \lambda \|\theta\|_1. \tag{3}$$

Identification assumptions, nuisance estimation, cross-fitting, and orthogonality are provided in Appendix B, and Appendix B.1 gives the conditions under which θ transfers from logged exposure changes to a shielding policy, with the same-period loss measured separately. The identification of θ is local to the budget-capped exposure changes of the logs, and the shielding sweep in Table 1, which removes up to 90% of a cohort, probes how far the fitted response extrapolates beyond them.

Given $\widehat{\theta}$, write $m_{a,i,t} = \widehat{\theta}^{\top} \phi_{a,i,t}$ for the gross marginal effect on $Y_{a,t+1}$ of one unit of exposure share of item i at its current profile. Because shares sum to one, the fit constrains only contrasts between items, and the level of $m_{a,i,t}$ is fixed by the profile, which has no intercept, so an item on which the unit records no behavior has $m_{a,i,t} = 0$, a natural zero (Appendix A). The unit-level contribution $c_{a,i,t} = \widehat{\theta}^{\top}(s_{a,i,t} - s_{a,i,t-1})$, which equals $X_{a,i,t}\, m_{a,i,t}$ for an item present in one period and $X_{a,i,t}\, m_{a,i,t} + b_{a,i,t-1}\, \widehat{\theta}^{\top}(\phi_{a,i,t} - \phi_{a,i,t-1})$ for an item present in both, is a signed predicted utility change, not a magnitude. It is positive when exposure moves toward an item that raises utility or away from one that lowers it, and negative in the two remaining cases (Table 6, Appendix A). Contributions, masses, and credits below are all expressed in the units of the utility $u_{a,t}$.

Settlement turns these contributions into item-level credit in four steps, carried out within each attribution unit and illustrated in Fig. 3 in Appendix A. First, an item is admitted in unit a only if the sign of $c_{a,i,t}$ agrees with the realized change $Y_{a,t+1}$. The gate $\mathbf{1}\{c_{a,i,t} Y_{a,t+1} > 0\}$ removes contributions whose predicted direction the outcome contradicts, which under substitution cannot be attributed to a single item. The unchanged items of a unit and the changed items that the gate rejects are not credited. They form the compensation base of that unit, the supply against which the credited change is measured, as the non-shielded supply is in a shielding experiment. Second, an admitted contribution enters with mass $|c_{a,i,t}|$ and is oriented by the reference score $\sigma_{a,i,t}$ of Eq. (7) in Appendix A, the current score $m_{a,i,t}$ when it is nonzero and otherwise the previous-period score $\widehat{\theta}^{\top} \phi_{a,i,t-1}$. The previous-period score uses the same $\widehat{\theta}$ and is zero if the item was absent. When the current score is zero, the contribution comes from the previous period, which supplies the reference sign. Credit carries $\operatorname{sign}(\sigma_{a,i,t})$, keeping the value of removed supply on the removed item.

Third, the mass is discounted by replaceability. If item i were shielded, its share would go to the compensation base, the supply that the policy leaves in place. The compensation ratio $r_{i,t} \in [0, 1]$ is the share of item i's gross marginal effect that the marginal effects of that base recover, both evaluated with the same $\widehat{\theta}$ (Appendix F.2). The net contribution is the gross contribution times $1 - r_{i,t}$. The same ratio is applied when item i gains share, with the base as the displaced supply, the matched-direction reading of Appendix F.2. The discount $w(r_{i,t}) = 1 - r_{i,t}$ scales the admitted contribution, and the sign gate before it and the unit budget after it complete the substitution correction. The range $r_{i,t} \in [0, 1]$ is a restriction that the settlement imposes, not a property of the ratio, and it keeps settled credit from exceeding gross attributed mass (Appendix J). An item whose substitutes are better than itself, the case of a negative shielding effect with a positive gross effect noted in Section 2.1, has a ratio above one. Such an item is settled at $r_{i,t} = 1$ and receives zero credit rather than negative credit. The logs do not realize the reallocation of an arbitrary item. In production, $r_{i,t}$ is a clipped linear function of substitute-density features of the compensation base, Eq. (70). Its parameters are calibrated so that the offline estimate of a shielded set, settled through the gate and the unit budgets, matches the net loss observed in earlier randomized shielding rounds, Eq. (71), with the procedure and set-level diagnostics in Appendix J. The compensation base under a policy is the supply that the policy leaves in place, so $r_{i,t}$ is a function of the shielded set. The settled credit of a set is therefore not linear in its size even though θ is linear (Appendix B.1). The discounted mass of item i in unit a is

$$q_{a,i,t} = w(r_{i,t})\ |c_{a,i,t}|\ \mathbf{1}\{c_{a,i,t} Y_{a,t+1} > 0\}, \qquad i \in \mathcal{C}_{a,t}. \tag{4}$$

Fourth, the masses of a unit are settled against the budget of Section 2.1, the realized change $|Y_{a,t+1}|$ of that unit. The magnitudes of the unit-level credits are the generalized Kullback–Leibler (KL) projection of $q_{a,\cdot,t}$ onto the budget set $\{R \geq 0 : \sum_i R_i \leq |Y_{a,t+1}|\}$, whose solution rescales the masses of a unit by one common factor and gives zero to an item rejected by the gate (Eq. (23), Appendix C). With the orientation of the second step, the unit-level credit is

$$R_{a,i,t} = \operatorname{sign}(\sigma_{a,i,t})\ \alpha^{\mathrm{cap}}_{a,t}\,|Y_{a,t+1}|\,\frac{q_{a,i,t}}{\sum_j q_{a,j,t}}, \qquad \alpha^{\mathrm{cap}}_{a,t} = \min\Big\{1, \frac{\sum_j q_{a,j,t}}{|Y_{a,t+1}|}\Big\}, \tag{5}$$

with $R_{a,i,t} = 0$ when the unit has no admitted mass. Equivalently $|R_{a,i,t}| = q_{a,i,t}\min\{1, |Y_{a,t+1}|/\sum_j q_{a,j,t}\}$, so masses that fit within the realized change are left unchanged, and masses that exceed it are scaled down to it. The capped attribution coefficient $\alpha^{\mathrm{cap}}_{a,t}$ is the fraction of the unit's budget that the admitted discounted masses fill, not the share of $Y_{a,t+1}$ that the response model explains (Appendix C). The settled credit of an item aggregates its unit-level credits,

$$R^{\mathrm{SCSV}}_{i,t} = \sum_{a:\, i \in \mathcal{C}_{a,t}} R_{a,i,t}. \tag{6}$$

Units whose utility declined are settled in the same way as units whose utility rose, so periods of decline are not discarded. The proof of Proposition 2.1 is provided in Appendix C.

Proposition 2.1 (Settlement properties)**.** *For every unit a and period t, the credits of Eq.* (5) *satisfy $|R_{a,i,t}| \leq q_{a,i,t}$ and $\sum_i |R_{a,i,t}| \leq |Y_{a,t+1}|$, every credited item has $\operatorname{sign}(R_{a,i,t}) = \operatorname{sign}(\sigma_{a,i,t})$, and whenever the unit has admitted mass, $|R_{a,i,t}| \geq |R_{a,j,t}|$ if and only if $q_{a,i,t} \geq q_{a,j,t}$. Consequently $\sum_{i\in\mathcal{I}_t} |R^{\mathrm{SCSV}}_{i,t}| \leq \sum_a |Y_{a,t+1}|$.*

Proposition 2.1 establishes algebraic properties of the settlement map only. Whether a settled credit tracks the shielding effect $V_{i,t}$ of Section 2.1 depends on the attribution, transfer, discount, and settlement steps above (Appendix B), and Section 3 checks this tracking at the cohort level.

2.3 Value-guided Positive Supply Intervention under Finite Attention

Positive supply intervention is an ex ante allocation problem because cold-start or under-exposed items must be selected before their ex post settled credit is observed. We therefore distill settled credit into an uncertainty-aware prior scorer and use its outputs to drive intervention, with details provided in Appendix D. Let $P_{i,t}$ denote the risk-adjusted reference score of Eq. (32) and $c_{i,t} > 0$ the reference exposure cost used for selection and density comparisons, distinct from current exposure $v_{i,t} \geq 0$. For items with positive reference exposure we use $c_{i,t} = v_{i,t}$, whereas newly uploaded items have $v_{i,t} = 0$ and a positive planned ramp-up cost $c_{i,t}$. The framework below is the reference formulation, stating what a value-guided policy should optimize and under which conditions comparisons of prior scores are valid. The deployed policy is a simpler gate and boost on the same prior (Section 4, Appendix K), and its effect is established by the online experiments, not by the bounds of this section. The intervention proceeds in two stages. *Supply selection* screens candidate eligibility using planned costs, and *exposure allocation* determines actual support inputs while accounting for displaced supply. Both stages use the same cap C^+_t. Selection consumes no exposure, and the reference costs are not per-item upper bounds on the later inputs $h_{i,t}$. Appendix Fig. 4 illustrates this two-stage process.

Supply selection. Given a candidate pool $\mathcal{A}_t$, intervention values $P_{i,t} \geq 0$, exposure costs $c_{i,t} > 0$, and budget C^+_t, selection is the 0–1 knapsack problem that maximizes the total intervention value of the admitted items subject to their total reference cost not exceeding C^+_t (Eq. (36) in Appendix E). We use value density $\rho_{i,t} = P_{i,t}/c_{i,t}$ as a scalable selection score and admit the longest prefix in descending order of $\rho_{i,t}$ whose total reference cost fits within C^+_t. Proposition E.1 in Appendix E shows that this prefix attains at least $(1-\eta)\,\mathrm{OPT}_t$, with OPT_t the optimal knapsack value, whenever no single reference cost exceeds a fraction η of the cap.

Exposure allocation. For the selected set, the reference framework chooses support inputs $h_{i,t} \geq 0$ under the same cap C^+_t and the joint inventory constraint $v_t + \delta v_t \geq 0$, where $\delta v_t = \sum_i h_{i,t}\, a^+_{i,t}$ is the joint exposure change and $a^+_{i,t} = e_i - \sum_j \omega^+_{ij,t} e_j$ the direction of the action on item i. Each action is scored by its value density net of the displacement-weighted densities of the items it draws exposure from, $\widehat{\tau}_{i,t} = \rho_{i,t} - \sum_j \omega^+_{ij,t}\rho_{j,t}$, and an action with $\widehat{\tau}_{i,t} \leq 0$ is not used, so the objective

$\sum_i h_{i,t}\widehat{\tau}_{i,t}$ accounts for supported items displacing one another. Appendix E states the resulting linear program. Proposition E.2 bounds the gap between the expected next-period utility change produced by jointly feasible support inputs and their score-based net gain $\sum_i h_{i,t}\widehat{\tau}_{i,t}$ by $\sum_i h_{i,t}\varepsilon_{i,t} + \frac{\Lambda}{2}\|\delta v_t\|^2$. The bound is conditional on the gradient of the expected utility being Λ-Lipschitz on the feasible segment and on each score-based net response $\widehat{\tau}_{i,t}$ lying within $\varepsilon_{i,t}$ of the true marginal net response of the action. Appendix F.2 proves the bound and derives $\varepsilon_{i,t}$ from prediction, settlement, and replacement-reference accuracy for items with positive reference exposure.

3 Experimental Results

Key details. *Posterior value* denotes the settled credit $R_{i,t}^{\mathrm{SCSV}}$, *prior score* denotes its serving-time prediction, and the *offline estimate* for a shielded set is the sum of its settled credits, settled with that set's compensation ratios (Eq. (69), Appendix J) and expressed in the utility units of the logs. Offline columns report the signed relative change $\widehat{\Delta}_{\mathrm{shield}} = -\widehat{L}/U_{\mathrm{ref}}$ (Eq. (72)), so an estimated loss appears as a negative percentage, and online columns report the platform's treated-minus-control relative lifts. Both are normalized within their own periods and compared under the cross-period stability condition of Appendix B.1. Every comparison with a randomized effect is made at the level of a set, the level at which shielding is randomized. The ratio $r_{i,t}$ is calibrated once on earlier shielding rounds disjoint from the experiments reported here (Appendix J), and no scale parameter is fitted to them, so agreement across comparisons that differ in cohort, metric, and scale is not a within-experiment fit. Placebo and falsification tests and non-causal sanity checks are reported in Appendices G.2 and G.3. **Model and training details** are provided in Appendix H.

3.1 Online shielding experiments versus offline posterior predictions

Open-reason. The reported measurement scope is full-page, covering the major traffic surfaces. Random shielding of the items of 20% of core productive authors produced an online difference-in-differences effect of -1.351% against the control arm, close to the offline estimate of -1.350%. Shielding the top 10% of items ranked by long-term retention value produced -1.25% online versus a conservative offline estimate of -0.95%. These experiments validate SCSV at the shielded-set estimand level. Under finite attention with strong substitution and compensation, raw item-attributed engagement cannot distinguish incremental utility from reallocated demand. The compensation-aware posterior value matches the realized full-page shielding loss within an absolute deviation of 0.30 pp.

Table 1: Online versus offline (SCSV) changes in total watch time on the 50k-view breakout cohort under random shielding.

Shielding ratio	Online	SCSV
90%	-8.85%	-8.31%
70%	-2.70%	-2.03%
50%	-1.40%	-0.97%
30%	-0.73%	-0.39%
10%	-0.35%	-0.08%

Cold-start breakout watch-time validation. We next validate the posterior value on cold-start breakout supply, using single-column total watch time within the 50k-view breakout cohort as the target metric, under random shielding at ratios from 10% to 90% (Table 1). All arms share the same cohort, item definition, and utility metric, and differ only in intervention intensity. The online losses are convex in the shielding ratio. Shielding 90% of the cohort removes 8.85% of total watch time, while shielding 10% removes 0.35%, a 25-fold drop in loss against a 9-fold reduction in shielded volume. The reported offline and online changes have the same negative sign, and their loss magnitudes follow the same ordering and qualitative convex profile. Scaling the 90% loss in proportion to shielded volume would predict a relative change near -0.98% at 10% shielding, about three times the observed -0.35%. The absolute deviation lies between 0.27 and 0.67 percentage points, the relative deviation falls from 77% at 10% shielding to about 6% at 90%, and every offline estimate is smaller in magnitude than the corresponding online loss. Appendix J traces the offline curvature to the compensation ratio falling as the shielded set grows, and the shortfall to the settlement only removing credit and to the same-period part of the watch-time response, which lies outside the next-period estimand (Appendix B.1). At the higher ratios two protocol differences contribute to the remaining gap. The two evaluations were five weeks apart, so the later offline settlement faced stronger compensation from a larger supply pool. The 0–24 hour cold-start window, which carries about 20% of the online effect, is not fully covered by the offline settlement logs (Appendix J). A

cross-check on live revenue, suppressing 5% of SA streamers, the platform's highest streamer tier, gives -1.83% online against an offline estimate of -1.64% (Appendix J).

3.2 Conventional causal estimators and bidirectional consistency

The agreement in Section 3.1 could in principle be reached by any deconfounded estimator. Table 2 compares inverse propensity scoring (IPS), a propensity-weighted value baseline, doubly robust estimation (DR), and SCSV against the randomized 20% shielding effect. The log-based rows share the observational logs, covariates $W_{a,t}$, temporal split, and reporting convention of Eq. (72) (Appendix G.1). The baselines use additive slope-based predictions, while SCSV aggregates settled credits. SCSV has the smallest absolute deviation and standard error. The standard errors of IPS, propensity-weighted value, and DR are 3.0, 2.7, and 1.5 times as large. The last row projects the window-averaged control-minus-shield treatment contrast through the logged $\widehat{\theta}$ with no settlement step, giving -1.300% (SE 0.16%), 0.051 pp from the randomized effect (SE 0.140 pp); this window-level diagnostic jointly reflects first-period transfer and within-window changes of the contrast and the loss, and Appendix G.1 relates it to the components of Proposition B.2. The DR estimate is also close to the randomized effect, but aggregate comparisons alone do not isolate error due to the linear attribution layer. This comparison evaluates aggregate loss prediction. SCSV additionally provides item-level credits for the intervention framework of Section 2.3.

Table 2: The 20% random-shielding effect on open reasons under conventional causal estimators and the response projection of Appendix B.1.

Estimator	Estimate	SE	Dev.
RCT (ground truth)	-1.351%	–	–
IPS	-1.28%	0.42%	0.071
Propensity-weighted value	-1.19%	0.38%	0.161
Doubly robust	-1.33%	0.21%	0.021
SCSV (ours)	-1.350%	0.14%	0.001
Response projection	-1.300%	0.16%	0.051

After the shielding comparisons, we test the same signal in the opposite direction (Table 10 in Appendix G.1). A coin-promotion strategy is applied to core productive authors and evaluated on the supply-side (B-side) content open-reason count used in offline settlement. The online intervention yields a system-level uplift of $+0.230\%$ ($p = 0.03$). The offline estimate is formed before the intervention from the settled credits of the treated authors' items. The treated authors account for 11.44% of baseline platform open reasons and their estimated open-reason lift is 1.83%, which gives approximately $+0.21\%$ at the platform level. The agreement between $+0.21\%$ and $+0.230\%$ shows that the settled credits aggregate to the platform-level effect in the promotion direction as they do in the shielding direction. Both comparisons are at the cohort level, and item-level value under promotion is tested by the production experiments of Section 4.

3.3 Prior scoring and compensation validation

Sections 3.1 and 3.2 validate the ex-post settlement, and this section checks the serving-time prior distilled from it (Appendix D), with full diagnostics in Appendix I. The prior preserves the ranking of the compensated settlement label, with Kendall's $\tau = 0.852$ for deep evidential regression (DER) (Amini et al., 2020) against $\tau = 0.793$ for evidential deep learning (EDL) (Sensoy et al., 2018) trained on a thresholded label (Fig. 5). This τ measures rank fidelity to the settled label, and whether that label tracks shielding value is what Section 3.1 validates. Under the monotone eligibility gate, admitting 7.49% of items covers 29.07% of open-reason endpoints, and shielding 13% of supply chosen by the compensation-aware score yields a 2.167% drop on open reasons against 1.924% for 22% chosen by an uncompensated baseline (Appendix J). The value of the supply selected by the prior is reported in Table 3.

4 Large-Scale Production Deployment and Experimental Results

After posterior validation and prior intervention readiness are established, we deploy SCSV to evaluate value-guided positive supply work in production. A nearline service outputs a prior value score and an uncertainty signal for new uploads, and the online ranker admits items for ramp-up when their prior score exceeds the serving threshold and their uncertainty remains below the risk threshold. Admitted items are then boosted through the existing multi-factor ramp-up mechanism,

which receives the prior value score as an additional ranking feature. Implementation details and production diagnostics are provided in Appendix K and Appendix L, respectively.

Selection efficiency on lifetime supply value. The production evidence begins with whether SCSV selects supply with higher lifetime value than existing production metrics. Table 3 compares SCSV against the incumbent posterior lifetime-value baseline across cold-start view time, production open reasons, and live-streaming revenue. Across all three scenarios, SCSV identifies supply with stronger system-level impact. For open reasons, the row reports the shielding comparison of Section 3.3, in which SCSV selects 13% of supply and the incumbent 22%, as loss per unit of shielded supply. The $+90.3\%$ gain holds under these selection rates and is not a fixed-budget comparison.

Table 3: Supply-selection performance across three objectives. The open-reason row reports shielding-loss magnitude divided by the selected supply fraction.

Value Type	Incumbent	SCSV	Efficiency Gain
View Time	$+0.034\%$	$+0.111\%$	$+226\%$
Open Reasons	$+8.73\%$	$+16.62\%$	$+90.3\%$
Revenue	$+5.18\%$	$+7.61\%$	$+46.9\%$

End-to-end positive supply work. Table 4 reports weekly value gains from the online intervention, estimated by difference-in-differences (DiD), together with the positive-work ROI before and after SCSV deployment. The ROI divides the realized system-level gain on the corresponding metric by the attributed supply-side lift of the intervened supply, so it measures the rate at which supply-side work converts into system-level user value. View time increases by 0.053% ($p = 0.003$), core production open reasons by 0.363% ($p = 0.019$), and long-term retention by 0.037% ($p = 0.021$). The revenue estimate is positive ($+0.025\%$) but not statistically significant ($p = 0.074$). The ROI columns of Table 4 rise on every metric, so SCSV raises the marginal return of positive supply work in addition to its scale. Placebo and falsification tests for these DiD estimates, including pre-period trends, permutation inference, and negative-control outcomes, are reported in Appendix G.2. Additional pathway diagnostics in Appendix L show that the gains concentrate in active-intent open reasons such as share and revisit rather than in passive consumption.

Table 4: Weekly relative lifts from the user-randomized online difference-in-differences (DiD) experiment and positive-work ROI before and after SCSV deployment (full inference in Table 13). †The revenue lift is not distinguishable from zero at the 5% level ($p = 0.074$).

Value Type	Weekly Gain	ROI Before	ROI After	ROI Gain
View Time	$+0.053\%$	0.35%	1.37%	$3.9\times$
Open Reasons	$+0.363\%$	13.99%	50.00%	$3.6\times$
Retention	$+0.037\%$	0.07%	0.103%	$1.5\times$
Revenue	$+0.025\%$	0.23%	0.96%	$4.2\times^{\dagger}$

Substitution efficiency analysis. The production gains are consistent with the central SCSV mechanism, under which value-guided intervention allocates finite ramp-up traffic toward supply whose marginal contribution is difficult to compensate by substitutes. The positive-work ROI is the reciprocal of the substitution coefficient of Appendix L, so the $3.6\times$ open-reason ROI gain is the ratio of the incumbent boosting coefficient (7.16) to the SCSV coefficient (1.99), with a random-boosting reference (5.27) between them.

5 Conclusion

We presented SCSV, which combines lagged differencing, linear DML attribution, and compensation-aware settlement to construct from production logs a substitution-corrected credit for the utility lost when an item is shielded. The value signal guides positive supply intervention toward candidates with positive substitution-corrected marginal gain. Against randomized shielding ground truth, the settled credit agrees in sign and ordering with online losses at the level of shielded sets, within 0.3 pp on open reasons and with dose-dependent gaps on watch time. Deployment improves selection efficiency and positive-work ROI. These results support substitution-corrected marginal utility as a more reliable long-term objective than gross item-attributed engagement. Two limitations remain. The settled credit is an operational quantity validated on shielded sets, not an item-level consistent estimator of the shielding effect, and its scale follows realized unit-level utility changes. The lagged design targets the next-period response, leaves the same-period loss to direct measurement, and may miss longer-horizon substitution and delayed feedback.

AI USE STATEMENT

In this work, we used generative AI tools only for polishing author-written text, specifically for grammar and spelling correction and for improving the wording, sentence structure, and logical flow of the exposition. We have not used generative AI tools for research ideation, methodology or experimental design, theoretical derivation and proof, code writing, data analysis, result interpretation, or literature search and citation generation, and the remaining required disclosure tasks are not applicable to this work. We have reviewed all AI-assisted work: every suggested edit was checked by the authors to confirm that it did not change the technical meaning of the original text, and suggestions that did were rejected or rewritten. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Strategic adaptation. The framework does not model how creators respond once SCSV informs exposure allocation, and the two layers of the pipeline face different risks. The ex-post settlement leaves little room for direct manipulation, because an item's credit is capped by the utility changes realized in the units it is attributed in and discounted by replaceability. Engagement inflated without a realized utility change therefore settles no credit under the accounting the score is built on, and the durable way to raise one's score is to supply low-replaceability content that raises realized utility, the behavior the mechanism is intended to reward. The serving-time prior, in contrast, is a learned scorer over upload-time features and inherits the manipulation risks of any such model, for example mimicking scarcity signals to inflate the replaceability discount. Re-settlement against realized outcomes in each cycle, together with occasional randomized shielding, provides corrective anchors but no formal guarantee, and modeling collective adaptation, which would shift the substitution structure itself, is left to future work.

Broader impacts. A substitution-corrected valuation signal can improve ecosystem efficiency, but it can also be used to intensify engagement optimization in ways that amplify concentration, reduce exposure diversity, or disadvantage specific creator groups if not audited carefully. Because the framework is built on user-behavior logs, privacy-preserving aggregation, access controls, and careful measurement governance are important.

Safeguards and release considerations. Any public release of the proposed benchmark should be preceded by privacy review, de-identification or aggregation of user traces, and documentation of access restrictions, licenses, and terms of use. In deployment, uncertainty-aware gating, conservative traffic budgets, and post-deployment auditing are important safeguards against over-exposure of poorly identified or potentially harmful inventory. These safeguards have already been applied to the released minimal reproducible package, and will be re-applied to the full release upon acceptance.

REPRODUCIBILITY STATEMENT

The supplementary material accompanying this submission provides an initial reproducibility package, including (i) precise data-construction rules and split definitions (*data_construction_and_evaluation.md*), (ii) hyperparameters and training configuration (*hyperparameters_and_training.md*), (iii) hardware, runtime, and compute accounting (*hardware_and_compute.md*), (iv) anonymized pseudocode for the core estimation and settlement pipeline (*pseudocode_core_pipeline.py*), and (v) a de-identified data subset package (*dataset_small_package*) containing the field-level documentation, and a representative subset of anonymized records. Items (i) to (iv) are summarized in Appendix H. The full benchmark contains tens of millions of records. Benchmark V1, to be released upon acceptance, contains anonymized item-level modeling inputs and predictions, field-level schemas, fixed temporal splits, executable evaluation code, and aggregate reference outcomes from multiple randomized shielding and promotion experiments, including shielding ground-truth labels where disclosure is permitted, together with documentation of the anonymization and compliance procedure. The public benchmark supports reproduction of the task definition, the data-construction pipeline, the evaluation protocol, and the offline validation results of Section 3 without access to internal logs. The production results of Section 4 are computed on internal full-scale logs and are not reproducible from the public release.

REFERENCES

Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *Advances in neural information processing systems*, 33:14927–14937, 2020.

Guy Aridor, Duarte Gonçalves, Daniel Kluver, Ruoyan Kong, and Joseph Konstan. The economics of recommender systems: Evidence from a field experiment on movielens. Technical report, CESifo Working Paper, 2022.

Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.

Omer Ben-Porat, Itay Rosenberg, and Moshe Tennenholtz. Content provider dynamics and coordination in recommendation ecosystems. *Advances in Neural Information Processing Systems*, 33: 18931–18941, 2020.

Asia J. Biega, Krishna P. Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. In *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 405–414, 2018. doi: 10.1145/3209978.3210063.

Thomas Blake and Dominic Coey. Why marketplace experimentation is harder than it seems: The role of test-control interference. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, pp. 567–582, 2014.

Craig Boutilier, Martin Mladenov, and Guy Tennenholtz. Recommender ecosystems: A mechanism design perspective on holistic modeling and optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 22575–22583, 2024.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Chongming Gao, Shijun Li, Wenqiang Lei, Jiawei Chen, Biao Li, Peng Jiang, Xiangnan He, Jiaxin Mao, and Tat-Seng Chua. Kuairec: A fully-observed dataset and insights for evaluating recommender systems. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, pp. 540–550, 2022a.

Chongming Gao, Shijun Li, Yuan Zhang, Jiawei Chen, Biao Li, Wenqiang Lei, Peng Jiang, and Xiangnan He. Kuairand: An unbiased sequential recommendation dataset with randomly exposed videos. In *Proceedings of the 31st ACM international conference on information & knowledge management*, pp. 3953–3957, 2022b.

Negar Hazrati and Francesco Ricci. Choice models and recommender systems effects on users' choices. *User Modeling and User-Adapted Interaction*, 2024.

Zuwang He, Shihao Shu, Yuli Qu, Hanyu Gao, Ziliang Zhang, Diwei Chen, Xiangda Yan, Buyu Gao, Tanchao Zhu, Yumeng Li, and Junxiong Zhu. Canniuplift: A holistic framework for mitigating seller and incentive cannibalization in e-commerce uplift modeling. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2026. arXiv:2607.05242.

Keisuke Hirano and Guido W Imbens. The propensity score with continuous treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, pp. 73–84. Wiley, 2004.

Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the american statistical association*, 103(482):832–842, 2008.

Eugene Ie, Chih-wei Hsu, Martin Mladenov, Vihan Jain, Sanmit Narvekar, Jing Wang, Rui Wu, and Craig Boutilier. Recsim: A configurable simulation platform for recommender systems. *arXiv preprint arXiv:1909.04847*, 2019.

Guido W Imbens and Donald B Rubin. *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge university press, 2015.

Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. Unbiased learning-to-rank with biased feedback. In *Proceedings of the tenth ACM international conference on web search and data mining*, pp. 781–789, 2017.

Ramesh Johari, Hannah Li, Inessa Liskovich, and Gabriel Y Weintraub. Experimental design in two-sided platforms: An analysis of bias. *Management Science*, 68(10):7069–7089, 2022.

Edward H Kennedy, Zongming Ma, Matthew D McHugh, and Dylan S Small. Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society: Series B*, 79(4):1229–1245, 2017.

Ron Kohavi, Diane Tang, and Ya Xu. *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press, 2020.

Damien Lefortier, Adith Swaminathan, Xiaotao Gu, Thorsten Joachims, and Maarten de Rijke. Large-scale validation of counterfactual learning methods: A test-bed. *arXiv preprint arXiv:1612.00367*, 2016.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pp. 661–670, 2010.

Lihong Li, Wei Chu, John Langford, and Xuanhui Wang. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 297–306, 2011.

Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. A graph-based approach for mitigating multi-sided exposure bias in recommender systems. *ACM Transactions on Information Systems*, 40(2):1–31, 2022. doi: 10.1145/3470948.

Kelong Mao, Jieming Zhu, Liangcai Su, Guohao Cai, Yuru Li, and Zhenhua Dong. Finalmlp: An enhanced two-stream mlp model for ctr prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 4552–4560, 2023.

Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness and satisfaction in recommendation systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 2243–2251, 2018. doi: 10.1145/3269206.3272027.

Martin Mladenov, Elliot Creager, Omer Ben-Porat, Kevin Swersky, Richard Zemel, and Craig Boutilier. Optimizing long-term social welfare in recommender systems: A constrained matching approach. In *International Conference on Machine Learning*, pp. 6987–6998. PMLR, 2020.

Gourab K. Patro, Arpita Biswas, Niloy Ganguly, Krishna P. Gummadi, and Abhijnan Chakraborty. Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms. In *Proceedings of The Web Conference 2020*, pp. 1194–1204, 2020. doi: 10.1145/3366423.3380196.

James M Robins, Miguel A Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

Guillaume Saint-Jacques, Maneesh Varshney, Jeremy Simpson, and Ya Xu. Using ego-clusters to measure network effects at linkedin. *arXiv preprint arXiv:1903.08755*, 2019.

Yuta Saito, Shunsuke Aihara, Megumi Matsutani, and Yusuke Narita. Open bandit dataset and pipeline: Towards realistic and reproducible off-policy evaluation. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021.

Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*, pp. 1670–1679. PMLR, 2016.

Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/a981f2b708044d6fb4a71a1463242520-Paper.pdf.

Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2219–2228, 2018. doi: 10.1145/3219819.3220088.

Adith Swaminathan and Thorsten Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 814–823, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/swaminathan15.html.

Hanzhao Wang, Xiaocheng Li, and Kalyan Talluri. Transformer choice net: A transformer neural network for choice prediction. *arXiv preprint arXiv:2310.08716*, 2023.

Ruoxi Wang, Rakesh Shivanna, Derek Z. Cheng, Sagar Jain, Dong Lin, Lichan Hong, and Ed H. Chi. Dcn v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems. In *Proceedings of the Web Conference 2021*, pp. 1785–1797, 2021.

Fan Yao, Chuanhao Li, Denis Nekipelov, Hongning Wang, and Haifeng Xu. How bad is top-k recommendation under competing content creators? In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 39674–39701. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/yao23b.html.

Kevin Zielnicki, Guy Aridor, Aurélien Bibaut, Allen Tran, Winston Chou, and Nathan Kallus. The value of personalized recommendations: Evidence from netflix. *arXiv preprint arXiv:2511.07280*, 2025.

A NOTATION AND SIGN CONVENTIONS

Table 5: Notation used in Section 2 and Appendix E. The last four rows belong to the shielding comparisons of Appendices B.1 and G.1.

Symbol	Meaning
a, t, i	attribution unit, production period, item
C_t, $C_{a,t}$, C_t^+	surface exposure cap, attention volume of unit a, support-input cap
$v_{i,t}$	current exposure intensity of item i
$c_{i,t}$	positive reference exposure cost for selection and density comparisons
U_t, $U_{t+1}(v)$	ecosystem utility, utility of period $t+1$ when the period-t allocation is v
$\mathcal{H}_t$	information available when the period-t allocation is decided
$T_{-i}(v_t)$, $V_{i,t}$	allocation after shielding i under the serving policy, shielding effect Eq. (1)
$\lvert Y_{a,t+1}\rvert$, $\alpha_{a,t}^{\mathrm{cap}}$	settlement budget of unit a, capped attribution coefficient, Eq. (5)
$u_{a,t}$, $Y_{a,t+1}$	unit utility, lagged utility change $u_{a,t+1} - u_{a,t}$
$b_{a,i,t}$, $X_{a,i,t}$	exposure share of item i in unit a, lagged change $b_{a,i,t} - b_{a,i,t-1}$
$\mathcal{C}_{a,t}$	items with $s_{a,i,t} \neq s_{a,i,t-1}$
$s_{a,i,t}$, $\phi_{a,i,t}$, $\Phi_{a,t}$	behavior vector of unit a on item i, per-share profile $s_{a,i,t}/b_{a,i,t}$, aggregate treatment $\sum_{i\in\mathcal{C}_{a,t}}(s_{a,i,t} - s_{a,i,t-1})$
$W_{a,t}$	observed user, traffic, surface, and allocation confounders
θ, $m_{a,i,t}$	attribution weights, gross marginal utility per unit exposure share $\theta^\top \phi_{a,i,t}$
$\sigma_{a,i,t}$	reference score for credit orientation: current score if nonzero, previous-period score otherwise, Eq. (7)
$c_{a,i,t}$	signed unit-level contribution $\widehat{\theta}^\top(s_{a,i,t} - s_{a,i,t-1})$, equal to $X_{a,i,t}\, m_{a,i,t}$ for an item present in one period
$r_{i,t}$, $w(r_{i,t})$	compensation ratio, Eq. (70), replaceability discount $1 - r_{i,t}$
$q_{a,i,t}$	discounted sign-consistent mass of item i in unit a, Eq. (4)
$R_{a,i,t}$, $R_{i,t}^{\mathrm{SCSV}}$	unit-level credit, Eq. (5), and settled credit $\sum_a R_{a,i,t}$, Eq. (6)
$P_{i,t}$, $\rho_{i,t}$	risk-adjusted reference score; value density $P_{i,t}/c_{i,t}$
$\pi_{i,t}$, $u_{i,t}$	serving-time prior score and its uncertainty, Eq. (31)
$h_{i,t}$	nonnegative exposure input of the action supporting item i
$\omega_{ij,t}^+$	fraction of action i's exposure taken from item j
$a_{i,t}^+$	exposure direction $e_i - \omega_{i,t}^+$ of action i
δv_t	signed inventory change $\sum_i h_{i,t} a_{i,t}^+$
$\mathcal{D}_{i,t}$	displaced set $\{j : \omega_{ij,t}^+ > 0\}$
$\widehat{\tau}_{i,t}$, $\tau_{i,t}$	score-based and true marginal net responses along $a_{i,t}^+$
$\Delta R_{i,t}$, $\Delta I_{i,t}$, $M_{i,t}$	incremental return, displaced value, net gain $\Delta R_{i,t} - \Delta I_{i,t}$
$\widehat{M}_{i,t}$	score-based net gain, Eq. (38)
$X_{a,i,s}^{\mathrm{win}}$, $\Phi_{a,s}^{\mathrm{win}}$, $\widehat{d}_s^{\mathrm{win}}$	window contrast against the last pre-shielding period, Appendix G.1
U_{ref}, μ_s^0, $\bar{\mu}_{\mathrm{win}}^0$, λ_{win}	settlement-period total of the logs; control-arm utility of period s and its window mean; window relative loss, Eqs. (72) and (19)
ι, $W_{a,t}^-$, $\Delta f_t(z)$, η_t	value weights, $u_{a,t} = \iota^\top \sum_i s_{a,i,t}$; pre-period components of $W_{a,t}$; level term of the bridge; response-invariance residual, Appendix B.1
$L_s(z)$, $d_t(z)$	cohort loss of period s; treatment contrast, Eq. (16)

Table 6: Sign of $c_{a,i,t} = X_{a,i,t}\, m_{a,i,t}$, the contribution of an item present in one period, and the outcome direction it predicts. The gate in Eq. (4) admits an item when the realized $Y_{a,t+1}$ has the same sign as $c_{a,i,t}$, and nonzero credit carries the reference sign of Eq. (7), equal to $\mathrm{sign}(m_{a,i,t})$ in these single-period cases.

$m_{a,i,t}$	$X_{a,i,t}$	$c_{a,i,t}$	Consistent $Y_{a,t+1}$	Reading	Credit
>0	>0	>0	>0	more exposure to an item that raises utility	>0, gain
<0	<0	>0	>0	less exposure to an item that lowers utility	<0
<0	>0	<0	<0	more exposure to an item that lowers utility	<0
>0	<0	<0	<0	less exposure to an item that raises utility	>0, loss of removed supply

Table 7: Observability of the utilities that define supply value.

Quantity	Meaning	Availability
$U_{t+1}(v_t)$	utility with item i	logged
$U_{t+1}(v_t - v_{i,t}e_i)$	utility with i removed and no substitute	subtract i's own attributed utility, under additive attribution
$U_{t+1}(T_{-i}(v_t))$	utility with i removed and substitutes reallocating	not realized in logs, approximated by settlement

Because exposure shares sum to one within a unit, $\sum_i X_{a,i,t} = 0$. An intercept in $\phi_{a,i,t}$ would enter $\Phi_{a,t}$ multiplied by this sum and is not identified, so $\phi_{a,i,t}$ carries none, and the fit of Eq. (3) constrains only contrasts between the marginal effects of different items. The level of $m_{a,i,t}$ is set by the profile. An item on which the unit records no behavior, $\phi_{a,i,t} = 0$, has $m_{a,i,t} = 0$, so $m_{a,i,t}$ is a gross marginal effect measured from that natural zero. The gate and the sign of the credit depend on this level, whereas predictions of unit-level changes do not. This is one reason why the settled credit is validated against randomized shielding (Section 3) and is not taken as an estimate of $V_{i,t}$ by construction.

Sign gate. Units whose sign of $c_{a,i,t}$ contradicts $Y_{a,t+1}$ are excluded from settlement. Under substitution and leakage such a contradiction may reflect a substitute's reverse compensation, a same-period perturbation, or reallocation noise, none of which can be attributed to a single item. The gate decides eligibility only. Three properties of the pipeline limit its influence on settled credits, and two external checks bound what remains. Magnitudes come from $\widehat{\theta}$, which is fit across all units. The gate is symmetric, excluding $c > 0$ with $Y < 0$ and $c < 0$ with $Y > 0$ alike. The credit settled in a unit is capped by the change that unit realized, however many items are admitted. The serving-time prior is trained on settled credits without access to $Y_{a,t+1}$ (Appendix D). A label that merely rationalized realized outcomes would carry no more forward information than realized engagement itself, whereas priors trained on settled credits select supply of higher shielding-verified value than priors trained on uncompensated engagement labels (Section 3.3 and Table 3). The settled credits themselves are compared with randomized shielding in Section 3. The gate credits an item only through its own aligned contribution. Benefit that reaches the ecosystem through other items, for example by seeding their discovery, accrues to those items rather than to the source, so sources of indirect value are under-credited. This errs toward withholding rather than over-allocating ramp-up traffic. Multi-hop attribution is left to future work.

Allocation mechanism. The shielding map T_{-i} is specified by the serving policy (Section 2.1). In the setting studied here the released exposure is reallocated to the remaining inventory, and the interventions of Eq. (34) hold the current total exposure fixed, $\mathbf{1}^\top(v_t + \delta v_t) = \mathbf{1}^\top v_t \le C_t$. Within a unit the shares sum to one, and across units the ranker distributes a finite surface budget, so exposure added to one item is exposure removed from others. This is the mechanism behind cannibalization and compensation (Fig. 1). Equal displacement follows from this reallocation restriction, not from the capacity inequality $\sum_i v_{i,t} \le C_t$ alone, and conservation by itself places no sign restriction on how the values of sets combine (Appendix B.1). If the removed item's exposure were not reallocated, the shielding effect would reduce to the conditional expected gross utility difference $\mathbb{E}[U_{t+1}(v_t) - U_{t+1}(v_t - v_{i,t}e_i) \mid \mathcal{H}_t]$, corresponding to the gross value in Table 7. The treatment is a supply-side change in an item's share of this budget, decided by the ranker from item-level performance aggregated across users, and not a change in an individual user's tastes. Personalization is not excluded by this view. It determines who within unit a sees which items, and its unit-fixed component is removed by the differencing in Eq. (2) while its time-varying component enters through $W_{a,t}$. The treatment is item-side, and the attribution unit only indexes where the outcome is measured.

Net versus gross value. Three utilities enter the definition of supply value and differ in how observable they are (Table 7). The gross value $U_{t+1}(v_t) - U_{t+1}(v_t - v_{i,t}e_i)$ uses only the first two and is cheap under additive attribution but biased under substitution. The net value $V_{i,t}$ uses the third, which the settled credit of Section 2.2 approximates from logs through deconfounded increments and the compensation ratio $r_{i,t}$.

Table 8: Value objects. No separate item-level estimator $\widehat{V}_{i,t}$ is introduced, and the settled credit is compared with randomized shielding only through the set-level offline estimate (Section 3).

Object	Symbol	Meaning
Shielding effect	$V_{i,t}$	next-period utility lost when item i is removed and the serving policy reallocates its exposure, Eq. (1); the causal target, policy-dependent, may be negative
Settled credit	$R_{i,t}^{\mathrm{SCSV}}$	operational credit assigned ex post within each attribution unit by attribution, the sign gate, the replaceability discount, and the unit budget, then summed over units, Eq. (6); each unit-level credit is oriented by $\sigma_{a,i,t}$ (Eq. (7))
Serving-time prior	$\pi_{i,t}$, $u_{i,t}$	prediction of the settled credit from upload-time features and its uncertainty, Eq. (31); the reference framework uses the risk-adjusted score $P_{i,t} = [\pi_{i,t} - \kappa u_{i,t}]_+$, Eq. (32)
Offline loss estimate	$\widehat{L}(\mathcal{S})$	sum of settled credits over a shielded set $\mathcal{S}$, Eq. (69); reported as $\widehat{\Delta}_{\mathrm{shield}} = -\widehat{L}/U_{\mathrm{ref}}$ when compared with signed relative shielding effects

Estimand, settled credit, and prior. Table 8 lists the value objects used in this paper. They are related but not interchangeable, and only the first is a causal quantity. In the experiments the settled credit is also called the posterior value and the serving-time prior the prior score.

The settlement budget is not implied by the estimand. Consider a unit whose utility is 10 in periods t and $t+1$ because the loss from a reduction in the exposure of item i was offset by an unrelated gain. Then $Y_{a,t+1} = 0$, no item passes the gate in this unit, and the unit settles no credit, although removing item i lowered its utility. The budget is an accounting rule that caps what a unit settles by what it realized, not an identification result. The quantity $Y_{a,t+1}$ is a temporal change and can reflect demand shocks, seasonality, or supply changes alike, and the confounders $W_{a,t}$ enter the attribution step, not the budget. Two consequences follow. The cap acts within a unit by one common factor, so it preserves the within-unit order of the discounted masses and never raises a mass (Proposition 2.1). Across units the factor differs, so the settled credit of an item, and the offline estimate $\widehat{L}(\mathcal{S})$ built from it (Eq. (69)), inherit their scale from the realized unit-level changes and from $\widehat{\theta}$, and no scale parameter is fitted to randomized effects. Agreement in magnitude with randomized shielding is therefore an empirical question, examined in Section 3, and not a property guaranteed by the settlement rule.

Worked example. Consider one attribution unit, three items, and periods $t-1$, t, $t+1$. At $t-1$ item 1 has exposure share 1.0 and utility is $u_{t-1} = 10$. At t the ranker shifts share from item 1 to item 2 (b_1 goes from 1.0 to 0.4 and b_2 from 0 to 0.6), with item 1's per-share profile unchanged so that its term is its share change at its profile, and utility stays at $u_t = 10$. At $t+1$ utility falls to $u_{t+1} = 9$. Then $X_{1,t} = -0.6$, $X_{2,t} = +0.6$, and $Y_{t+1} = -1$. Three effects are separated. The first effect is confounding. A fixed taste of this unit for item 1 would raise both its exposure and its utility in every period, and is removed by the differences in Eq. (2). The second effect is interference. Removing 0.6 of item 1 while item 2 fills in costs one unit of utility, not item 1's full gross contribution, so the net drop already embeds item 2's compensation. The third effect is leakage. Item 2's high exposure and response at t are not used as its value, because the outcome window opens only after the exposure change is fixed. The sign gate then keeps item 1 ($c_1 < 0$, $Y < 0$) and drops item 2 ($c_2 > 0$, $Y < 0$). Item 1 is credited $+\min\{q_1, |Y|\}$ with $q_1 = (1 - r_1)\, 0.6\, m_1$, at most the one unit of utility that was lost, so the lost utility is settled onto the removed item rather than the item that momentarily gained exposure.

Reference score. The reference score that orients unit-level credit in Section 2.2 is

$$\sigma_{a,i,t} = \begin{cases} m_{a,i,t}, & m_{a,i,t} \neq 0, \\ \widehat{\theta}^{\top} \phi_{a,i,t-1}, & m_{a,i,t} = 0. \end{cases} \tag{7}$$

Orientation of credit. For positive discounted mass, the discount preserves the reference sign $\mathrm{sign}(\sigma_{a,i,t})$. Negative unit-level credits have $\sigma_{a,i,t} < 0$. The reference sign determines settlement orientation, not the sign of $V_{i,t}$. The gate and the orientation follow from the shielding semantics of

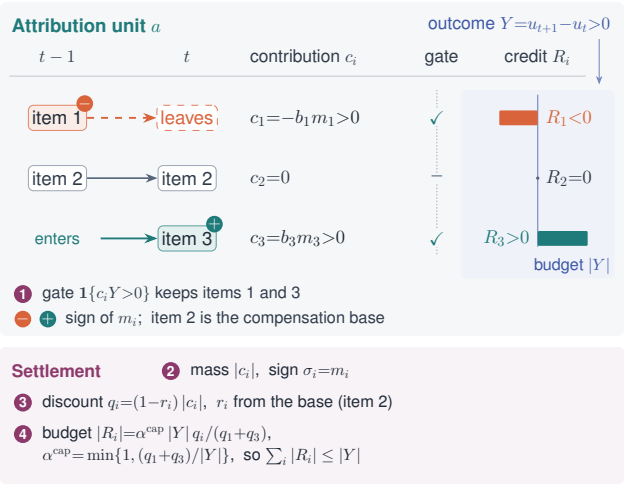


Figure 3: Settlement within one attribution unit, with the four steps of Section 2.2 numbered. A harmful item leaves and a helpful item enters, both pass the sign gate, and their discounted masses are settled against the unit budget $|Y|$ with the sign of the reference score, negative for the removed harmful item and positive for the added helpful item.

Section 2.1. They keep the loss of removed supply on the removed item, and the discount converts gross into net contribution within the range $r_{i,t} \in [0, 1]$ in which the two have the same sign.

B Deconfounded SCSV Estimation Details

This appendix details the residualization, cross-fitting, and orthogonality used to estimate the attribution weight vector θ in Eq. (3).

Setup. Recall the attribution-unit temporal differences in Section 2.2. For attribution unit a and period t, the pseudo-outcome is $Y_{a,t+1} = u_{a,t+1} - u_{a,t}$, and the sparse item-level treatment change is $X_{a,i,t} = b_{a,i,t} - b_{a,i,t-1}$, and the treatment term of item i is the change in its behavior vector $s_{a,i,t}$ (Section 2.2). The aggregated treatment feature is

$$\Phi_{a,t} = \sum_{i \in \mathcal{C}_{a,t}} (s_{a,i,t} - s_{a,i,t-1}), \tag{8}$$

where $\mathcal{C}_{a,t} = \{i : s_{a,i,t} \neq s_{a,i,t-1}\}$. For an item present in one period the term equals $X_{a,i,t}\phi_{a,i,t}$ with $\phi_{a,i,t}$ its per-share profile, and for an item present in both periods it equals $X_{a,i,t}\phi_{a,i,t} + b_{a,i,t-1}(\phi_{a,i,t} - \phi_{a,i,t-1})$, the share change at the current profile plus the change in per-share behavior at the previous share, which the implementation does not separate from it (Appendix H). For example, if the share of item 1 rises by 0.10 while the shares of items 2 and 3 fall by 0.06 and 0.04, with unchanged profiles, then $\Phi_{a,t} = 0.10\,\phi_1 - 0.06\,\phi_2 - 0.04\,\phi_3$. We posit the partially linear model

$$Y_{a,t+1} = \theta^\top \Phi_{a,t} + f(W_{a,t}) + \varepsilon_{a,t}, \tag{9}$$

where $W_{a,t}$ denotes observed user, traffic, surface, and system-allocation confounders.

Nuisance functions and residualization. Define the nuisance functions

$$m(W_{a,t}) = \mathbb{E}[\Phi_{a,t} \mid W_{a,t}], \qquad \ell(W_{a,t}) = \mathbb{E}[Y_{a,t+1} \mid W_{a,t}]. \tag{10}$$

Given estimates $\widehat{m}(\cdot)$ and $\widehat{\ell}(\cdot)$, form residuals

$$\widetilde{\Phi}_{a,t} = \Phi_{a,t} - \widehat{m}(W_{a,t}), \qquad \widetilde{Y}_{a,t+1} = Y_{a,t+1} - \widehat{\ell}(W_{a,t}). \tag{11}$$

Table 9: Sampling timeline of the settlement logs and of a shielding experiment. The logs give one sample $(\Phi_{a,t}, W_{a,t}, Y_{a,t+1})$ per unit and period. In the experiment the policy is in force on days $t, \ldots, t+H-1$, and the platform's difference-in-differences averages the daily arm differences over these days and subtracts their pre-period mean.

Period	Settlement logs, unit a	Shielding experiment, cohort
before $t-1$	lookback window of $W_{a,t}^{-}$: user, traffic, surface, and allocation covariates, earlier utility levels	pre-period; arms identical in expectation
$t-1$	$b_{a,i,t-1}$, $s_{a,i,t-1}$, $u_{a,t-1}$; pre-treatment	last pre-shielding day
t	$b_{a,i,t}$, $s_{a,i,t}$; treatment $\Phi_{a,t} = \sum_i (s_{a,i,t} - s_{a,i,t-1})$; level $u_{a,t} = u_{a,t-1} + \iota^\top \Phi_{a,t}$, included in $W_{a,t}$	first shielded day, first day of the window of Eq. (19); contrast $d_t(z)$; same-period loss $L_t(z) = \iota^\top d_t(z)$
$t+1$	$u_{a,t+1}$; outcome $Y_{a,t+1} = u_{a,t+1} - u_{a,t}$	second shielded day; $L_{t+1}(z)$ of Eq. (16): the loss of day t carried forward, the lagged response to $d_t(z)$, the level term, and the same-period loss of day $t+1$
$t+2, \ldots, t+H-1$	–	remaining shielded days; the window of Eq. (19) is $t, \ldots, t+H-1$
$t+H$	–	lagged response of the last shielded day, outside the platform's window

Cross-fitting. To reduce overfitting bias, we use K-fold cross-fitting over attribution-unit samples $\mathcal{D} = \{(a,t)\}$. Split $\mathcal{D}$ into folds $\{\mathcal{D}_k\}_{k=1}^K$. For each fold k, fit nuisance models $\widehat{m}^{(-k)}(\cdot)$ and $\widehat{\ell}^{(-k)}(\cdot)$ on $\mathcal{D} \setminus \mathcal{D}_k$, and compute residuals on $\mathcal{D}_k$:

$$\widetilde{\Phi}_{a,t} = \Phi_{a,t} - \widehat{m}^{(-k)}(W_{a,t}), \qquad \widetilde{Y}_{a,t+1} = Y_{a,t+1} - \widehat{\ell}^{(-k)}(W_{a,t}), \quad (a,t) \in \mathcal{D}_k. \tag{12}$$

Orthogonalized estimation. SCSV estimates θ using the orthogonalized regression in Eq. (3):

$$\widehat{\theta} \in \arg\min_{\theta} \sum_{a,t} \left(\widetilde{Y}_{a,t+1} - \theta^\top \widetilde{\Phi}_{a,t} \right)^2 + \lambda \|\theta\|_1. \tag{13}$$

The associated score is Neyman-orthogonal. First-order errors in nuisance estimation affect the target estimate only at higher order. This improves robustness when $W_{a,t}$ is high-dimensional and the nuisance models are imperfect.

Identification and consistency.

Assumption B.1 (Identification of θ)**.** (i) Conditional exogeneity of the change means $\mathbb{E}[\varepsilon_{a,t} \mid \Phi_{a,t}, W_{a,t}] = 0$, so that after differencing and conditioning on $W_{a,t}$ the treatment term, the change in the behavior the unit records on each item, holds no information about the utility innovation. (ii) The timing condition is that no component of $W_{a,t}$ measured before period t is affected by $X_{a,i,t}$, which holds by construction. Write $W_{a,t}^{-}$ for these components. The utility level $u_{a,t}$ is also included in $W_{a,t}$, so that mean reversion in the outcome is not attributed to $X_{a,i,t}$. It is contemporaneous with the treatment, $u_{a,t} - u_{a,t-1} = \iota^\top \Phi_{a,t}$ (Section 2.2), so conditioning on it fixes the value-weighted total of the period-t change, and θ is the next-period response to the composition of that change at a fixed period-t utility level. Condition (i) is required conditional on $u_{a,t}$. The component of $\Phi_{a,t}$ along ι is the change of the level itself. Once $u_{a,t}$ and $u_{a,t-1}$ are conditioned on, it has no residual variation, and its coefficient is not separated from the level dependence of f (the singular case of condition (e) in Appendix B.1). A same-period utility loss then enters the next period through the level itself and through the level dependence of f, which Appendix B.1 carries as separate terms. (iii) The residual variation condition is that $\mathrm{Var}(\Phi_{a,t} \mid W_{a,t})$ is non-degenerate on the composition subspace $\{\iota^\top \Phi = 0\}$, on which θ is identified.

Condition (i) is imposed on the change $X_{a,i,t}$ rather than on the level $b_{a,i,t}$. Unit-fixed determinants of exposure cancel in the difference, so the condition is weaker than conditional exogeneity of the level. Under Assumption B.1, the partially linear model of Eq. (9), cross-fitting, and nuisance estimators whose product error rate is $o(n^{-1/2})$, the unpenalized residual-on-residual estimator, Eq. (13) with $\lambda = 0$ and a fixed dimension of θ, is consistent and asymptotically normal by the results

of Chernozhukov et al. (2018). The ℓ_1 penalty is used for stability over the sparse active candidate universe and introduces shrinkage, so we do not rely on asymptotic normality of the penalized estimate, and the uncertainty reported in Table 2 comes from the cluster bootstrap of Appendix G.1. Condition (i) is not testable from logs. Appendix G.2 reports the placebo and falsification checks that probe it, and Section 3.1 reports agreement with randomized shielding, where assignment is independent of unobservables by design.

Implementation notes. The attribution design is sparse because only temporally changed items enter $\mathcal{C}_{a,t}$. We therefore solve Eq. (13) with scalable regularized linear solvers over an active candidate universe. The linear attribution form supports auditing, since $\widehat{\theta}^\top(s_{a,i,t} - s_{a,i,t-1})$ gives the deconfounded unit-level contribution used in Eq. (4). Any orthogonalized estimator of θ can replace the DML step, since settlement uses only the unit-level contributions $c_{a,i,t}$.

B.1 From the logged response to a shielding policy

Objects. Let z denote a shielding policy under which a set $\mathcal{S}$ of items is removed from exposure in period t and the ranker reallocates their shares, so $b_{a,i,t}(z) = 0$ for $i \in \mathcal{S}$ and the shares of the remaining items rise. Let $q \in \{0, z\}$ index the control and the policy, and write $\Phi_{a,t}(q)$, $u_{a,s}(q)$, and $Y_{a,t+1}(q)$ for the corresponding potential values of the quantities in Section 2.2. Over the attribution units covered by the experiment define

$$d_t(z) = \mathbb{E}[\Phi_{a,t}(0) - \Phi_{a,t}(z)]\,, \qquad L_s(z) = \mathbb{E}[u_{a,s}(0) - u_{a,s}(z)]\,. \tag{14}$$

$d_t(z)$ is the control-minus-shield contrast throughout. It includes the shielded items and the changes on substitutes, so compensation is already included. The policy-induced change $\mathbb{E}[\Phi_{a,t}(z) - \Phi_{a,t}(0)]$ has the opposite sign, $-d_t(z)$. It must not be confused with the temporal change $X_{a,i,t}$ within either arm. $L_{t+1}(z)$ is the next-period, cohort-level form of the shielding estimand of Section 2.1. It measures the utility lost when $\mathcal{S}$ is removed and attention reallocates. $L_t(z)$ is its same-period counterpart, which the estimand excludes. Write $W^-_{a,t}$ for the components of $W_{a,t}$ measured before period t and ι for the value weights of Section 2.2, $u_{a,t} = \iota^\top \sum_i s_{a,i,t}$. Table 9 aligns the sampling of the logs with the days of an experiment. Thus $L_s(z) > 0$ denotes a utility reduction, whereas the reported shielding change is negative after normalization by a positive reference utility (Eq. (72)).

Conditions. (a) Assumption B.1 holds in the logs. (b) Baseline invariance of the pre-period covariates. The pre-period covariates satisfy $W^-_{a,t}(0) = W^-_{a,t}(z)$, and the arms differ in $W_{a,t}$ only through the utility level $u_{a,t}$. The policy changes the shares $b_{a,i,t}$, hence the behavior of period t, so $u_{a,t}$ is not invariant. By the identity $u_{a,t} - u_{a,t-1} = \iota^\top \Phi_{a,t}$ and the invariance of period $t-1$, the same-period loss is $L_t(z) = \iota^\top d_t(z)$, the arm difference of the first shielded day, an observable quantity rather than an assumption. (b_0) Same-period invariance. The utility level satisfies $u_{a,t}(0) = u_{a,t}(z)$, so that $\iota^\top d_t(z) = 0$ and, with (b), $W_{a,t}(0) = W_{a,t}(z)$. That is, substitutes recover the value-weighted behavior of the shielded items within period t. This is the case in which condition (ii) of Assumption B.1 holds for the policy without qualification. (c) Response invariance. For the values of $\Phi_{a,t}(z)$ the policy produces,

$$\mathbb{E}[Y_{a,t+1}(q) \mid \Phi_{a,t}(q), W_{a,t}] = f(W_{a,t}) + \theta^\top \Phi_{a,t}(q), \qquad q \in \{0, z\}. \tag{15}$$

(d) Measurement consistency. The period length, the transform $\phi_{a,i,t}$, the utility $u_{a,t}$, and the unit a are the same in the logs and in the experiment. (e) If the residual Gram matrix $Q = \mathbb{E}[\widetilde{\Phi}_{a,t}\widetilde{\Phi}^\top_{a,t}]$ is singular, θ is identified only up to $\mathrm{Null}(Q)$, and $\theta^\top d_t(z)$ is identified exactly when $d_t(z) \in \mathrm{Range}(Q)$. Because $d_t(z)$ is a difference of means of $\Phi_{a,t}$, this fails only when the policy moves $\Phi_{a,t}$ in a direction the logs never vary. The direction ι is the case in point. When $u_{a,t}$ and $u_{a,t-1}$ both enter $W_{a,t}$, $\iota^\top \Phi_{a,t}$ has no residual variation, θ is identified on the composition subspace $\{\iota^\top \Phi = 0\}$, and its coefficient along ι is not separated from the level dependence of f. A policy with $\iota^\top d_t(z) \neq 0$ moves $\Phi_{a,t}$ along exactly this direction, so its projection $\theta^\top d_t(z)$ is identified only under (b_0). Proposition B.2 carries the unidentified part in two explicit terms.

Proposition B.2 (Response bridge)**.** *Under (b) and (d), let η_t be the difference between the arms in $\mathbb{E}[Y_{a,t+1}(q) - f(W_{a,t}(q)) - \theta^\top \Phi_{a,t}(q)]$, which is zero under (c), and let $\Delta f_t(z) = \mathbb{E}[f(W_{a,t}(0)) - f(W_{a,t}(z))]$. Then*

$$L_{t+1}(z) = L_t(z) + \theta^\top d_t(z) + \Delta f_t(z) + \eta_t, \qquad L_t(z) = \iota^\top d_t(z). \tag{16}$$

The first term is the same-period loss, which the next-period level inherits. The second is the lagged response to the treatment contrast, and the third is the change in f induced by the arm difference in $u_{a,t}$. Under (b_0), $L_t(z) = 0$ and $\Delta f_t(z) = 0$. Under (b_0) and (c), $L_{t+1}(z) = \theta^\top d_t(z)$.

Proof. By the definition of η_t and iterated expectations, $\mathbb{E}[Y_{a,t+1}(0) - Y_{a,t+1}(z)] = \theta^\top d_t(z) + \Delta f_t(z) + \eta_t$, and by Eq. (2) the left side equals $L_{t+1}(z) - L_t(z)$. Since period $t-1$ is common to both arms under (b), $L_t(z) = \mathbb{E}[u_{a,t}(0) - u_{a,t}(z)] = \iota^\top \mathbb{E}[\Phi_{a,t}(0) - \Phi_{a,t}(z)] = \iota^\top d_t(z)$. Under (b) the arms differ in $W_{a,t}$ only through $u_{a,t}$, so $\Delta f_t(z)$ is the arm difference of f over the same-period loss. Under (b_0), $W_{a,t}(0) = W_{a,t}(z)$, so $L_t(z) = 0$ and $\Delta f_t(z) = 0$. Under (c), $\eta_t = 0$. □

The result is a transfer statement, not a new estimator. Its content is condition (c), which states that the response of next-period utility to the treatment representation, fitted on the budget-capped increments of the logs, is the one the policy produces. Eq. (16) separates the next-period loss into four terms. The first is the same-period loss $L_t(z)$, observable on the first shielded day. The second is the lagged response $\theta^\top d_t(z)$. The third is the level term $\Delta f_t(z)$. The fourth is η_t, which under a policy that stays in force on period $t+1$ contains the same-period loss of that period, because the removal is still in place and the model of Eq. (9) takes no period-$(t+1)$ allocation as input. The fit of Appendix B conditions on $u_{a,t}$ and identifies θ on the composition subspace of condition (e). When $\iota^\top d_t(z) \neq 0$, the component of $\theta^\top d_t(z)$ along ι and $\Delta f_t(z)$ are both consequences of the same-period loss. The fit does not separate them, and their sum is not the slope of $\mathbb{E}[u_{a,t+1} \mid u_{a,t}, u_{a,t-1}, W^-_{a,t}]$ in $u_{a,t}$ either, because conditioning on the level does not hold the composition of $\Phi_{a,t}$ fixed. We do not estimate these two terms. The daily series of an experiment gives $L_t(z)$ and $L_{t+1}(z)$ directly, and under (b_0) both terms vanish. Appendix G.1 relates the window comparison of Table 2 to these channels. Table 2 uses $\widehat{\theta}$ and experimental treatment contrasts for a window-level diagnostic. Its relation to Eq. (16) requires the time and reference alignment specified below (Appendix G.1).

Compensation. Let g_z denote the gross counterfactual of Section 2.1 at the cohort level, in which $\mathcal{S}$ is removed and its exposure is not reallocated. Writing $d_t(z) = d_t^{\text{gross}} + d_t^{\text{rec}}$ with $d_t^{\text{gross}} = \mathbb{E}[\Phi_{a,t}(0) - \Phi_{a,t}(g_z)]$ and $d_t^{\text{rec}} = \mathbb{E}[\Phi_{a,t}(g_z) - \Phi_{a,t}(z)]$, the projection of Eq. (16), which equals $L_{t+1}(z)$ under (b_0) and (c), becomes

$$\theta^\top d_t(z) = A_\theta - H_\theta = A_\theta\,(1 - r_\theta), \qquad A_\theta = \theta^\top d_t^{\text{gross}}, \quad H_\theta = -\theta^\top d_t^{\text{rec}}, \quad r_\theta = H_\theta / A_\theta, \tag{17}$$

for $A_\theta \neq 0$. A_θ is the gross loss of the removed supply and H_θ the utility that substitutes recover. This is the cohort-level form of the discount $w(r_{i,t}) = 1 - r_{i,t}$ in Section 2.2. Writing $A_\theta = \sum_{i \in \mathcal{S}} A_i$ with A_i the gross term of item i, the recovery collects the absorbing supply's terms that each removed share produces, $H_\theta = \sum_{i \in \mathcal{S}} A_i\, r_{i,t}$, so r_θ is the gross-weighted average of the item ratios under the policy's reallocation. Nothing in Eq. (17) confines r_θ to $[0, 1]$. When $A_\theta > 0$, $r_\theta > 1$ is the over-compensated case in which the shielded set is replaced by better supply and $L_{t+1}(z) < 0$. For $A_\theta < 0$, the sign of $A_\theta(1 - r_\theta)$ must be evaluated with that negative gross term. Table 1 reports positive observed loss magnitudes and positive settled loss estimates $\widehat{L}$, both displayed as negative changes. This agreement in sign does not identify the gross term A_θ or the compensation ratio. Under (b_0) and (c), a positive next-period loss implies that A_θ and $1 - r_\theta$ have the same sign. If $A_\theta > 0$ is established separately, then $r_\theta < 1$. These validation rounds are separate from the earlier calibration rounds, which are not reported. The item-level restriction $r_{i,t} \in [0, 1]$ of Section 2.2 is imposed by the settlement, not identified by the experiments. In settlement the contribution $c_{a,i,t}$ of a shielded item is its own term $X_{a,i,t}\, m_{a,i,t}$, hence gross, while the recovery sits in the substitutes' terms, which the sign gate removes when $Y_{a,t+1} < 0$. The discount restores that recovery at the item level. The projection $\theta^\top d_t(z)$ is already net and is not discounted again. Neither arm realizes g_z, so the experiment identifies only the net loss, never A_θ or H_θ separately. Under (b_0) and (c) the net loss is the difference $A_\theta - H_\theta$, and the calibration of $r_{i,t}$ targets the net loss (Appendix J).

Dose. For a policy $z = (\mathcal{S}, p)$ that shields a fraction p of a cohort, the reallocation depends on what remains, so $d_t(\mathcal{S}, p)$ need not be linear in p. A linear θ is therefore compatible with the convex loss curve of Table 1. Likewise $\theta^\top[\, d_t(\mathcal{S} \cup \mathcal{T}) - d_t(\mathcal{S}) - d_t(\mathcal{T})\,]$ carries no sign restriction, so the losses of disjoint sets need not add. Holding the compensation ratios and the other settlement inputs fixed, the expected aggregate value under random shielding is linear in the marginal inclusion probability. Evaluating the compensation ratios on the remaining supply instead allows set-dependent values. The

value of a set then need not equal the sum of separately evaluated subset values, and the dose profile follows from the set-dependent discounts and the unit-level caps. The offline estimates in Table 1 use this set-dependent construction and follow the same qualitative convex profile as the online losses.

Error decomposition. If (c) holds only approximately, η_t of Proposition B.2 is nonzero. Under (b_0), so that $L_{t+1}(z) = \theta^\top d_t(z) + \eta_t$, and with $\widehat{L}_P = \widehat{\theta}^\top \widehat{d}_t$,

$$\widehat{L}_P - L_{t+1}(z) = (\widehat{\theta} - \theta)^\top d_t(z) + \widehat{\theta}^\top \big(\widehat{d}_t - d_t(z)\big) - \eta_t, \tag{18}$$

and a settled estimate $\widehat{L}^{\mathrm{settle}}$, normalized to the same per-unit scale, metric scope, and period as $\widehat{L}_P$, adds the term $\widehat{L}^{\mathrm{settle}} - \widehat{L}_P$, which collects the gate, the discount, the budget, and the aggregation. A raw settled total cannot be subtracted from a per-unit projection. This decomposition is on the first-period loss scale. Reported changes have the opposite sign after division by the matched positive reference utility. The projection and settlement rows of Table 2 additionally require the window and normalization alignment described below and in Appendix G.1. A single gap does not attribute error among these terms. Without (b_0) the terms $L_t(z)$ and $\Delta f_t(z)$ of Eq. (16) enter as well, together with the component of $\theta^\top d_t(z)$ along ι that the log fit does not identify. The daily series of an experiment gives $L_t(z)$ directly, since the first shielded day carries only the same-period loss, but does not separate the remaining terms.

Window. The experiment holds the shielding policy fixed over H periods $t, \ldots, t+H-1$, and the platform measures the effect on these H shielded days. Under the lag convention the treatment of day s is paired with the outcome of day $s+1$, so the lagged responses to the window's contrasts fall on $t+1, \ldots, t+H$, one period later. Appendix G.1 states how this offset enters the projection diagnostic. Write $\mu_s^0 = \mathbb{E}[u_{a,s}(0)] > 0$, $\bar{\mu}^0_{\mathrm{win}} = H^{-1} \sum_{s=t}^{t+H-1} \mu_s^0$, and $\bar{L}_{\mathrm{win}} = H^{-1} \sum_{s=t}^{t+H-1} L_s(z)$, both over the shielded days. The window relative loss and its signed change are

$$\lambda_{\mathrm{win}} = \frac{\bar{L}_{\mathrm{win}}}{\bar{\mu}^0_{\mathrm{win}}} = \frac{\sum_{s=t}^{t+H-1} L_s(z)}{\sum_{s=t}^{t+H-1} \mu_s^0}, \qquad \Delta_{\mathrm{win}} = -\lambda_{\mathrm{win}}. \tag{19}$$

The platform's difference-in-differences relative lift estimates Δ_{win} over these same days. It averages the daily treated-minus-control relative differences over the window and subtracts their pre-period mean, which has zero expectation under random assignment. This ratio of window means is a control-utility-weighted average of the daily relative losses, and it need not equal their unweighted average. All window quantities use the same period weights. Equal daily weights are displayed here. Proposition B.2 covers the transition from the first shielded period to the next. In later periods the same-period loss recurs, the level difference persists, and $W_{a,s}$ differs between arms, so Eq. (16) cannot simply be summed. Write $\lambda_s = L_s(z)/\mu_s^0$ for the relative loss of day s. The additional condition (f), *within-window relative-loss stationarity*, states that λ_s is constant over the shielded days after the first, $s = t+1, \ldots, t+H-1$, the days on which a lagged response is present. Under (f),

$$\lambda_{\mathrm{win}} = \lambda_{t+1} + (\lambda_t - \lambda_{t+1}) \, \frac{\mu_t^0}{H \bar{\mu}^0_{\mathrm{win}}}, \tag{20}$$

so the window relative loss is the next-period relative loss of Proposition B.2 plus a first-day correction of weight about $1/H$. The correction vanishes when the first day's relative loss equals the later ones, and under (b_0), where $\lambda_t = 0$, it scales the window loss by $1 - \mu_t^0/(H\bar{\mu}^0_{\mathrm{win}})$. By Eq. (16), under (c) $\lambda_{t+1} = [L_t(z) + \theta^\top d_t(z) + \Delta f_t(z)]/\mu_{t+1}^0$, and under ($b_0$) as well $\lambda_{t+1} = \theta^\top d_t(z)/\mu_{t+1}^0$, even if the control utility changes over time. Condition (f) is a property of the daily series of an experiment, which shows the relative loss of the first shielded day, same-period by construction, and its evolution over the window. The comparisons of Appendix G.1 use the window average and maintain (f) as a condition. Condition (f) does not constrain the projected treatment contrast in later periods. Appendix G.1 retains that separate source of discrepancy.

Scale-invariant comparison across periods. We measure supply value as the fraction of ecosystem utility lost after removal and reallocation. For an evaluation period r, write this relative loss as $\lambda_r = L_r/B_r$, where loss L_r and positive baseline utility B_r use the same metric scope and aggregation. The representation satisfies $(c_r L_r)/(c_r B_r) = \lambda_r$ for any $c_r > 0$, so proportional growth of baseline utility and shielding loss preserves the relative value of the removed supply. This gives a common comparison target across periods without fixing the ecosystem's absolute utility level.

The log-based estimate $\widehat{\lambda}_{\log} = \widehat{L}/U_{\text{ref}}$ uses its own settlement-period reference (Appendix J). Write $\lambda_{\log}$ for the corresponding true relative loss. For a matched metric scope, evaluated supply, and policy dose, condition (f′), *cross-period relative-loss stability*, states $\lambda_{\log} = \lambda_{\text{win}}$. Condition (f) connects the next-period loss of the bridge to the window, while (f′) connects the log period to the experimental window. Under (f′), the independently normalized log and experimental effects target the same relative loss, with no cross-period baseline conversion. More generally, their discrepancy has the exact decomposition

$$\widehat{\Delta}_{\text{shield}} - \Delta_{\text{win}} = -\big(\widehat{\lambda}_{\log} - \lambda_{\log}\big) - \big(\lambda_{\log} - \lambda_{\text{win}}\big). \tag{21}$$

The first term measures prediction error within the log period. The second measures changes in relative supply value across periods and is zero under (f′). Comparison to the estimated RCT effect adds experimental estimation error. This decomposition distinguishes a change in ecosystem scale from a change in substitution or compensation. The former preserves the ratio when loss grows proportionally, whereas the latter can change the ratio itself. In the watch-time evaluation, the five-week separation allows the substitution environment to evolve, and the incomplete 0–24 hour log coverage introduces a further scope difference. The observed deviations therefore measure the full offline-to-online gap, including these documented protocol differences. The response bridge, the settlement, and the cross-period comparison address distinct links in this construction. More generally, the results rely on platform-specific logs, interventions, and utility metrics, so external validity may be limited where inventory dynamics or ranking constraints differ.

C Proof of Proposition 2.1 (Settlement properties)

Recall the settlement rule in Section 2.2. For unit a and period t write $Q = \sum_j q_{a,j,t} \geq 0$ and $B = |Y_{a,t+1}|$. If $Q = 0$ every credit of the unit is zero and all claims hold trivially. For any $q_{a,i,t} > 0$, the gate implies $c_{a,i,t} \neq 0$ and $B > 0$. If $m_{a,i,t} \neq 0$, then $\sigma_{a,i,t} = m_{a,i,t} \neq 0$. Otherwise, $c_{a,i,t} = -\widehat{\theta}^\top s_{a,i,t-1} = -b_{a,i,t-1}\sigma_{a,i,t} \neq 0$, so again $\sigma_{a,i,t} \neq 0$. Thus every positive mass has a nonzero reference score, and Eq. (5) gives

$$|R_{a,i,t}| = \lambda_a \, q_{a,i,t}, \qquad \lambda_a = \min\{1, B/Q\} \in (0, 1]. \tag{22}$$

The magnitudes of the unit-level credits are the generalized KL projection of the discounted masses onto the budget set,

$$\arg\min_{R \geq 0} \sum_{i \in \mathcal{C}_{a,t}} \Big(R_i \log \tfrac{R_i}{q_{a,i,t}} - R_i + q_{a,i,t} \Big) \quad \text{s.t.} \quad \sum_{i \in \mathcal{C}_{a,t}} R_i \leq |Y_{a,t+1}|, \tag{23}$$

under the conventions $0 \log(0/q) = 0$ for every $q \geq 0$ and $R \log(R/0) = +\infty$ for $R > 0$, so that an item rejected by the gate, $q_{a,i,t} = 0$, receives $R_i = 0$. The solution is a common rescaling within the unit.

KL projection. The objective of Eq. (23) is convex and separable. Items with $q_{a,i,t} = 0$ have $R_i = 0$ under the convention stated with Eq. (23). For admitted items, $q_{a,i,t} > 0$, with a multiplier $\mu \geq 0$ for the budget constraint, stationarity gives $\log(R_i/q_{a,i,t}) + \mu = 0$, that is $R_i = q_{a,i,t} e^{-\mu}$. If $Q \leq B$ the constraint is slack, $\mu = 0$, and $R_i = q_{a,i,t}$. If $Q > B$ the constraint binds and $e^{-\mu} = B/Q$. Both cases give $R_i = \lambda_a q_{a,i,t}$, the magnitude in Eq. (22).

Properties. Since $\lambda_a \leq 1$, $|R_{a,i,t}| \leq q_{a,i,t}$. Summing over items, $\sum_i |R_{a,i,t}| = \lambda_a Q = \min\{Q, B\} \leq |Y_{a,t+1}|$. Every nonzero $R_{a,i,t}$ has the sign of $\sigma_{a,i,t}$ by Eq. (5). Because $\lambda_a > 0$ multiplies every mass of the unit, $|R_{a,i,t}| \geq |R_{a,j,t}|$ if and only if $q_{a,i,t} \geq q_{a,j,t}$. Finally, by the triangle inequality, $\sum_i |R^{\text{SCSV}}_{i,t}| \leq \sum_a \sum_i |R_{a,i,t}| \leq \sum_a |Y_{a,t+1}|$. □

Reading of the capped coefficient. A unit whose admitted mass reaches its realized change settles the whole change, and a unit whose admitted mass falls short of it settles only that mass and leaves the rest of the change unsettled. It is not the share of $Y_{a,t+1}$ that the response model explains. The total $\sum_j q_{a,j,t}$ is a same-sign sum taken after the gate and the discount, so contributions that cancel in $\widehat{\theta}^\top \Phi_{a,t}$ do not reduce it. The realized change $Y_{a,t+1}$ also includes the part that the nuisance $f(W_{a,t})$ accounts for, which the budget does not separate (Appendix A).

D UNCERTAINTY-AWARE PRIOR SCORING FOR COLD-START INTERVENTION

This appendix details the uncertainty-aware prior scorer used for online cold-start intervention. The ex-post settled credit in Eq. (6) is available only after next-period utility, and with it the unit budgets, is observed. Online intervention, however, must decide whether to allocate exposure before this ex-post value is realized. We therefore distill settled credits into a serving-time prior score and attach an epistemic uncertainty estimate for conservative exposure allocation.

Training target and serving-time features. For item i in period t, let

$$y_{i,t} = T\big(R_{i,t}^{\mathrm{SCSV}}\big) \tag{24}$$

denote the supervised target, where $T(\cdot)$ is an optional monotone normalization or variance-stabilizing transform used for training. Let $z_{i,t}$ denote features available at serving time, including content features, creator features, historical creator behavior, lightweight metadata, and substitution-related context. Importantly, $z_{i,t}$ excludes future outcomes used to construct $R_{i,t}^{\mathrm{SCSV}}$, so the prior model can be applied before ex-post settlement is available.

Evidential regression model. The prior scorer maps serving-time features $z_{i,t}$ to Normal-Inverse-Gamma parameters

$$(\gamma_{i,t}, \nu_{i,t}, \alpha_{i,t}, \beta_{i,t}) = h_\psi(z_{i,t}), \tag{25}$$

with positivity constraints enforced by

$$\begin{aligned} \nu_{i,t} &= \mathrm{softplus}(\widetilde{\nu}_{i,t}) + \epsilon_\nu, \\ \alpha_{i,t} &= 1 + \mathrm{softplus}(\widetilde{\alpha}_{i,t}) + \epsilon_\alpha, \\ \beta_{i,t} &= \mathrm{softplus}(\widetilde{\beta}_{i,t}) + \epsilon_\beta, \\ \gamma_{i,t} &\in \mathbb{R}, \end{aligned} \tag{26}$$

where $\epsilon_\nu, \epsilon_\alpha, \epsilon_\beta > 0$ are small numerical constants. The Normal-Inverse-Gamma parameterization represents uncertainty over both the predictive mean and the observation variance.

Marginalizing the latent mean and variance yields a Student-t predictive distribution:

$$p(y_{i,t} \mid z_{i,t}) = \mathrm{St}\left(y_{i,t}; \gamma_{i,t}, \frac{\beta_{i,t}(1+\nu_{i,t})}{\nu_{i,t}\alpha_{i,t}}, 2\alpha_{i,t}\right), \tag{27}$$

where $\gamma_{i,t}$ is the predictive location, $\frac{\beta_{i,t}(1+\nu_{i,t})}{\nu_{i,t}\alpha_{i,t}}$ is the predictive scale parameter, and $2\alpha_{i,t}$ is the degree of freedom.

We use

$$\mu_{i,t} = \mathbb{E}[y_{i,t} \mid z_{i,t}] = \gamma_{i,t} \tag{28}$$

as the prior value on the training-target scale, and use the epistemic uncertainty proxy

$$\sigma_{i,t}^{\mathrm{epi}} = \sqrt{\frac{\beta_{i,t}}{\nu_{i,t}(\alpha_{i,t}-1)}} \tag{29}$$

to quantify uncertainty in the predictive mean. A larger $\sigma_{i,t}^{\mathrm{epi}}$ indicates that the model has weaker evidence for the prior value estimate, which is common for cold-start or distribution-shifted supply.

Training objective. The evidential prior scorer is trained by minimizing the Student-t negative log-likelihood with an evidence regularizer:

$$\mathcal{L}_{i,t}^{\mathrm{evi}} = -\log p(y_{i,t} \mid z_{i,t}) + \lambda_{\mathrm{ev}} \left|y_{i,t} - \gamma_{i,t}\right| (2\nu_{i,t} + \alpha_{i,t}), \tag{30}$$

where $\lambda_{\mathrm{ev}} > 0$ controls the strength of evidence regularization. The regularizer penalizes high evidence when prediction error is large, discouraging over-confident estimates on poorly identified or shifted supply.

Prior score and uncertainty interface. At serving time, the model produces a prior value score and an uncertainty score:

$$\pi_{i,t} = \mu_{i,t}, \qquad u_{i,t} = \sigma_{i,t}^{\mathrm{epi}}. \tag{31}$$

If the training target is normalized by $T(\cdot)$, the score is mapped back to the serving scale using the corresponding inverse calibration, and the uncertainty is carried through the same map so that value and uncertainty are expressed on one scale. To first order $u_{i,t} = |(T^{-1})'|\, \sigma_{i,t}^{\mathrm{epi}}$, and for nonlinear T this transfer is asymmetric, so $u_{i,t}$ summarizes it symmetrically. For the reference selection and allocation framework of Section 2.3, define the risk-adjusted reference score

$$P_{i,t} = [\pi_{i,t} - \kappa u_{i,t}]_+, \tag{32}$$

where $\kappa \geq 0$ is a user-specified uncertainty-penalty coefficient, held fixed across candidates within a decision instance and dimensionless because $\pi_{i,t}$ and $u_{i,t}$ share a scale. Setting $\kappa = 0$ recovers the nonnegative prior score $[\pi_{i,t}]_+$. Larger values impose a stronger penalty. This construction does not assert a calibrated confidence bound. The deployed rule instead applies separate thresholds to prior value and uncertainty, which is not equivalent to thresholding $\pi_{i,t} - \kappa u_{i,t}$. Serving uses the following monotone eligibility gate

$$\mathrm{eligible}(i,t) = \mathbf{1}\{\pi_{i,t} \geq \tau_s \ \wedge \ u_{i,t} \leq \upsilon_s\}, \tag{33}$$

where τ_s and υ_s are surface-specific thresholds calibrated on validation data and online experiments.

Production instantiation. We train an upload-time prior model that maps serving-available item, creator, and recent behavioral features to a prior value score π_i and an uncertainty proxy u_i. The supervision target is the compensated SCSV settlement label, while an uncompensated label, denoted as "origin", is used as a baseline for ablation. The feature set follows production availability constraints and includes content-quality signals, creator descriptors, recent creator behavior aggregates, lightweight metadata, and substitution-related context. At serving time, the intervention policy consumes (π_i, u_i) through a monotone joint eligibility gate on prior value and uncertainty. The key evaluation criterion is therefore whether the learned prior preserves the ranking of substitution-corrected posterior value, beyond pointwise prediction loss.

Connection to value-guided intervention. The uncertainty-aware prior score does not modify the ex-post SCSV settlement rule in Eq. (6). Its properties (Proposition 2.1) hold by construction. Instead, uncertainty modeling provides a serving-time approximation to the settled credit before future outcomes are observed. In Section 2.3, the intervention policy uses $P_{i,t}$ as the value input for supply selection and exposure allocation, discounting predicted value by its uncertainty for high-variance cold-start supply while preserving the substitution-corrected valuation principle learned from settled SCSV labels.

E Value-guided Intervention: Objective, Exposure Allocation, and Illustration

Intervention objective. This appendix states in full the intervention objective of Section 2.1 and the exposure-allocation stage of Section 2.3, and illustrates the two-stage process. For a candidate set $\mathcal{A}_t$, let $h_{i,t} \geq 0$ be the exposure assigned by the action supporting item i. Fix displacement weights $\omega_{ij,t}^{+} \geq 0$ over the inventory, with $\omega_{ii,t}^{+} = 0$ and $\sum_j \omega_{ij,t}^{+} = 1$. This action takes $h_{i,t}\omega_{ij,t}^{+}$ exposure from item j. Define its direction and the joint net change by

$$a_{i,t}^{+} = e_i - \omega_{i,t}^{+}, \qquad \delta v_t = \sum_{i \in \mathcal{A}_t} h_{i,t} a_{i,t}^{+}, \qquad \mathbf{1}^\top \delta v_t = 0, \tag{34}$$

where e_i is the coordinate vector of item i and $\omega_{i,t}^{+}$ collects its displacement weights. The intervention reallocates existing exposure under the allocation mechanism of Appendix A, preserving $\sum_j v_{j,t}$ within the surface cap C_t. The separate cap C_t^{+} limits the gross exposure assigned by the support actions. The platform seeks to improve next-period utility through jointly feasible actions:

$$\begin{aligned} \max_{h \geq 0} \quad & \mathbb{E}[U_{t+1}(v_t + \delta v_t) - U_{t+1}(v_t) \mid \mathcal{H}_t], \\ \text{s.t.} \quad & \sum_{i \in \mathcal{A}_t} h_{i,t} \leq C_t^{+}, \qquad v_t + \delta v_t \geq 0. \end{aligned} \tag{35}$$

The joint nonnegativity constraint accounts for support actions drawing on the same substitute inventory. An item's final exposure change can differ from its support input because other actions can displace it. For a single action, the relevant gain is $M_{i,t}(h_{i,t}) = \Delta R_{i,t}(h_{i,t}) - \Delta I_{i,t}(h_{i,t})$, where $\Delta R_{i,t}$ is the incremental return from supporting item i and $\Delta I_{i,t}$ is the utility loss induced by its displaced supply. Section 2.3 constructs a score-based approximation to the joint utility objective.

Supply selection. Given a candidate pool $\mathcal{A}_t$, intervention value $P_{i,t} \geq 0$, exposure cost $c_{i,t} > 0$, and budget C_t^+, selecting supply solves

$$\max_{x_i \in \{0,1\}} \sum_{i \in \mathcal{A}_t} x_i P_{i,t}, \qquad \text{s.t.} \quad \sum_{i \in \mathcal{A}_t} x_i c_{i,t} \leq C_t^+. \tag{36}$$

This is a 0–1 knapsack problem. We use value density $\rho_{i,t} = \frac{P_{i,t}}{c_{i,t}}$ as a scalable selection score and admit the longest prefix in descending order of $\rho_{i,t}$ whose total reference cost fits within C_t^+.

Proposition E.1 (Near-optimal density selection)**.** *Assume $P_{i,t} \geq 0$, $c_{i,t} > 0$, and $\frac{\max_{i \in \mathcal{A}_t} c_{i,t}}{C_t^+} \leq \eta$. Let $\mathcal{S}_t^{\text{den}}$ be the density-ordered prefix selected by $\rho_{i,t}$. Then $\sum_{i \in \mathcal{S}_t^{\text{den}}} P_{i,t} \geq (1-\eta)\,\text{OPT}_t$, where OPT_t is the optimal value of Eq.* (36)*.* The proof is provided in Appendix F.

Exposure allocation. For the selected set $\mathcal{S}_t^{\text{den}}$, use the fixed displacement directions of Eq. (34), with $h_{i,t} = 0$ outside that set. Define $\mathcal{D}_{i,t} = \{j : \omega_{ij,t}^+ > 0\}$. The exposure displaced by a single support action satisfies the conservation identity

$$\sum_{j \in \mathcal{D}_{i,t}} h_{i,t} \omega_{ij,t}^+ = h_{i,t}, \tag{37}$$

which follows from $\sum_j \omega_{ij,t}^+ = 1$. Using the reference costs and densities of all supported and displaced items, the score-based net gain is

$$\widehat{M}_{i,t}(h_{i,t}) = h_{i,t} \widehat{\tau}_{i,t}, \qquad \widehat{\tau}_{i,t} = \rho_{i,t} - \sum_{j \in \mathcal{D}_{i,t}} \omega_{ij,t}^+ \rho_{j,t}. \tag{38}$$

Allocation maximizes the joint score-based gain over feasible support inputs:

$$\begin{aligned} \max_{h \geq 0} \quad & \sum_{i \in \mathcal{S}_t^{\text{den}}} h_{i,t} \widehat{\tau}_{i,t}, \\ \text{s.t.} \quad & \sum_{i \in \mathcal{S}_t^{\text{den}}} h_{i,t} \leq C_t^+, \qquad v_t + \delta v_t \geq 0, \\ & h_{i,t} = 0 \quad \text{for } i \in \mathcal{S}_t^{\text{den}} \text{ with } \widehat{\tau}_{i,t} \leq 0. \end{aligned} \tag{39}$$

The zero decision is feasible. The joint inventory constraint accounts for all simultaneous withdrawals, including actions that displace other supported items. The net change of item j is $\delta v_{j,t} = h_{j,t} - \sum_i h_{i,t} \omega_{ij,t}^+$. Since $\sum_i h_{i,t} \widehat{\tau}_{i,t} = \rho_t^\top \delta v_t$, the objective already includes these displacements. With fixed weights and densities this is a linear program over a polyhedral feasible set. It supplies a local first-order objective. The following proposition, which Section 2.3 states in words, relates it to actual utility through valuation error and the quadratic remainder in the joint change δv_t.

Proposition E.2 (Value-guided net-gain approximation)**.** *Write $\bar{U}_{t+1}(v) = \mathbb{E}[U_{t+1}(v) \mid \mathcal{H}_t]$. Fix jointly feasible support inputs $h_{i,t}$ with joint change δv_t and score densities defined for every supported and displaced item. If $\nabla \bar{U}_{t+1}$ is Λ-Lipschitz on the feasible segment and $|\widehat{\tau}_{i,t} - \tau_{i,t}| \leq \varepsilon_{i,t}$ for the true marginal net responses $\tau_{i,t}$ of Appendix F.2, then*

$$\left| \bar{U}_{t+1}(v_t + \delta v_t) - \bar{U}_{t+1}(v_t) - \sum_i h_{i,t} \widehat{\tau}_{i,t} \right| \leq \sum_i h_{i,t} \varepsilon_{i,t} + \frac{\Lambda}{2} \|\delta v_t\|^2. \tag{40}$$

Appendix F.2 proves this bound. Appendix F derives $\varepsilon_{i,t}$ from prediction, settlement, and replacement-reference accuracy for items with positive reference exposure, otherwise $\varepsilon_{i,t}$ is a separate assumption. Deployment uses the prior-based gate of Appendix K, and Section 4 evaluates the realized gains.

Illustration. Fig. 4 summarizes the two-stage intervention of Section 2.3, suppressing time subscripts. Candidates are ranked by $\rho_i = P_i/c_i$. A support action of size h_i takes $h_i\omega_{ij}^+$ exposure from each $j \in \mathcal{D}_i$, giving a score-based return $h_i\rho_i$, displacement cost $h_i \sum_j \omega_{ij}^+\rho_j$, and net gain $h_i\widehat{\tau}_i$.

Selection. In the upper panel each bar is a candidate and its height is the reference exposure cost c_i. The longest density-ordered prefix whose total cost fits within C^+ is admitted. Proposition E.1 gives the guarantee $(1-\eta)\,\mathrm{OPT}$, with η bounding the largest individual cost as a fraction of the cap. This stage checks eligibility and planned-cost feasibility and consumes no exposure.

Allocation. The lower panel shows one support action, which adds h_i to item i and withdraws $h_i\omega_{ij}^+$ from each displaced item. Their sum equals h_i by Eq. (37). Simultaneous actions produce $\delta v = \sum_i h_i(e_i - \omega_i^+)$ and must jointly satisfy $v + \delta v \geq 0$. Allocation chooses actual inputs under the same cap C^+, with zero inputs for nonpositive $\widehat{\tau}_i$ as in Eq. (39). Proposition E.2 bounds the gap between the joint score-based gain and expected utility gain. Deployment uses the empirical prior-and-uncertainty gate of Appendix K. Its effect is evaluated in Section 4.

Correspondence with serving. The figure uses the quantities of Section 2.3, and Appendix K gives their served counterparts. $P_{i,t}$ is the risk-adjusted reference score $[\pi_{i,t} - \kappa u_{i,t}]_+$ of Eq. (32), so within the reference framework a cold-start item competes with a value reduced by its uncertainty. In deployment the gate $u_{i,t} \leq \upsilon_s$ instead withholds ramp-up eligibility from items whose prior is too uncertain, without removing them from regular recommendation. C_t^+ is the per-surface ramp-up cap. Within that cap, the served policy admits candidates through the joint gate and boosts them with the prior value score inside the existing ranker, and the eligibility threshold τ_s is an empirically calibrated admission threshold that is not derived from the density comparison of Eq. (38). Proposition E.1 bounds the density-ordered selector for additive rewards and fixed costs. The served policy admits items through the uncertainty gate and boosts them inside the existing ranker rather than sorting by the value density $P_{i,t}/c_{i,t}$. The rewards of boosted items also need not be additive, because boosted items can displace one another. The bound is therefore a property of the selector and not of the served policy, whose effect is supported by Table 4 and the production diagnostics of Section 4. Under matched small exposure changes against the same substitutes, shielding and boosting share the same first-order net response for supply that already has exposure. Appendix F connects this response to the settled-credit density by separating prediction and settlement discrepancies, finite-removal curvature, and replacement-reference differences, and Proposition E.2 bounds the gap between the score-based allocation objective and the expected ecosystem gain for jointly feasible increments. The settled credits from which the boosting prior is distilled are validated at the level of shielded sets by the shielding experiments of Section 3, with $r_{i,t}$ calibrated on earlier rounds (Appendix J). Newly uploaded supply has no exposure to remove, so the prior carries this reading over from items with exposure history, and its effect on such supply rests on the experiments of Section 4.

F Proofs for Value-guided Positive Supply Intervention

F.1 Proof of Proposition E.1

Sort candidates in descending order of value density $\rho_{i,t} = P_{i,t}/c_{i,t}$, and suppress the time index for clarity. Let $(1), (2), \ldots$ denote the sorted order, with $\rho_{(1)} \geq \rho_{(2)} \geq \cdots$. For $\eta \geq 1$ the claimed lower bound is nonpositive and the result is immediate. If all candidates fit, the prefix selects all of them and is optimal. Otherwise take $0 \leq \eta < 1$ and let K be the largest index such that

$$\sum_{k=1}^{K} c_{(k)} \leq C^+, \qquad \sum_{k=1}^{K+1} c_{(k)} > C^+. \tag{41}$$

The density-ordered prefix is $\mathcal{S}^{\mathrm{den}} = \{(1), \ldots, (K)\}$, with total value

$$P_{\mathrm{den}} = \sum_{k=1}^{K} P_{(k)}. \tag{42}$$

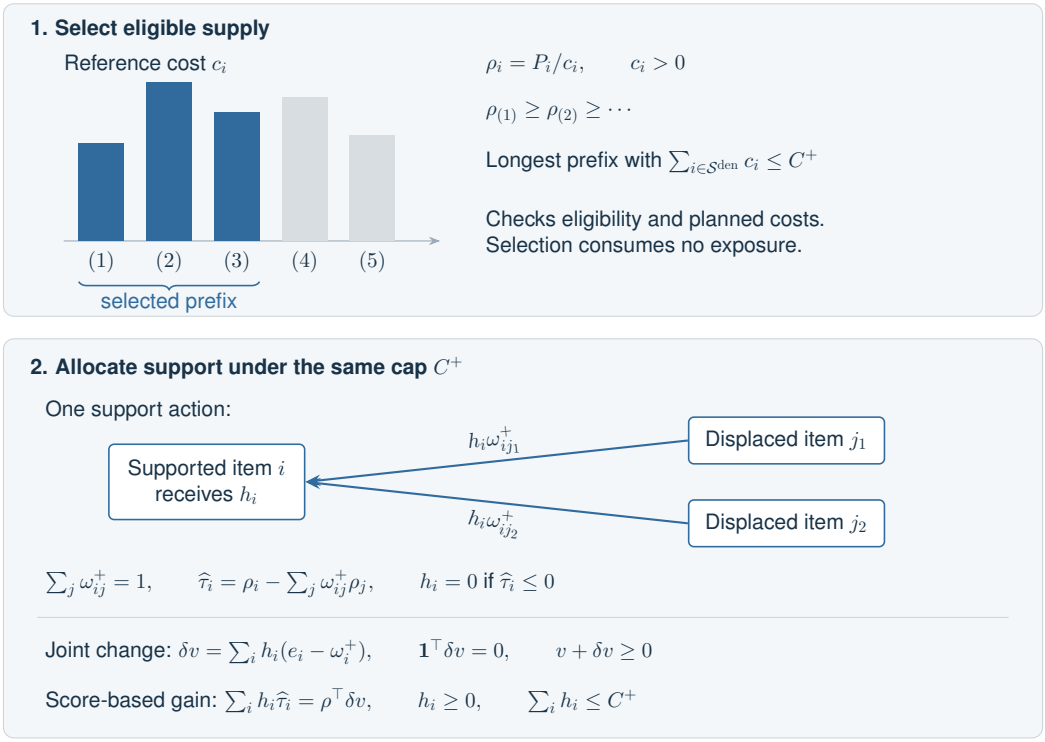


Figure 4: Value-guided selection and allocation, with time subscripts suppressed. Selection ranks candidates by $\rho_i = P_i/c_i$ and checks their reference costs against C^+. Allocation chooses $h_i \geq 0$ under the same cap, withdrawing $h_i\omega^+_{ij}$ from each displaced item. The joint change $\delta v = \sum_i h_i(e_i - \omega^+_i)$ preserves total exposure and satisfies $v + \delta v \geq 0$. The score-based net gain is $\sum_i h_i \widehat{\tau}_i = \rho^\top \delta v$, and Proposition E.2 relates it to expected utility gain.

Consider the fractional relaxation of Eq. (36), where $x_i \in [0, 1]$. The density-ordered fractional solution is optimal for this relaxation. Its value is

$$\text{OPT}_{\text{frac}} = P_{\text{den}} + \alpha P_{(K+1)}, \qquad \alpha = \frac{C^+ - \sum_{k=1}^{K} c_{(k)}}{c_{(K+1)}} \in [0, 1]. \tag{43}$$

Since the integer optimum is upper bounded by the fractional optimum,

$$\text{OPT} \leq \text{OPT}_{\text{frac}}. \tag{44}$$

By the small-item condition, $c_{(K+1)} \leq \eta C^+$. Moreover, because the fractional solution fills the budget using items with density at least $\rho_{(K+1)}$, we have

$$\text{OPT}_{\text{frac}} \geq \rho_{(K+1)} C^+. \tag{45}$$

Therefore,

$$P_{(K+1)} = \rho_{(K+1)} c_{(K+1)} \leq \eta \rho_{(K+1)} C^+ \leq \eta\, \text{OPT}_{\text{frac}}. \tag{46}$$

Using Eq. (43),

$$\begin{aligned} P_{\text{den}} &= \text{OPT}_{\text{frac}} - \alpha P_{(K+1)} && (47)\\ &\geq \text{OPT}_{\text{frac}} - P_{(K+1)} && (48)\\ &\geq (1-\eta)\text{OPT}_{\text{frac}} && (49)\\ &\geq (1-\eta)\text{OPT}. && (50) \end{aligned}$$

This proves

$$\sum_{i \in \mathcal{S}^{\text{den}}} P_i \geq (1-\eta)\text{OPT}. \tag{51}$$

□

F.2 Proof of Proposition E.2 and the valuation-to-allocation bridge

We suppress item-level time subscripts and fix the history, outcome window, and utility scale. We write $\bar{U}_{t+1}(v) = \mathbb{E}[U_{t+1}(v) \mid \mathcal{H}_t]$ for the conditional expected ecosystem utility of Section 2.1, and $\partial_j \bar{U}_{t+1}(v)$ for its marginal component in the exposure of item j. Every exposure direction below sums to zero, so a common additive shift of the marginal components cancels from every contrast, which is the contrast-only identification noted in Appendix A.

Quantitative statement. Restate the directions and joint feasibility of Section 2.3. Fix a selected set and weights $\omega_{ij}^{+} \geq 0$ with $\omega_{ii}^{+} = 0$ and $\sum_j \omega_{ij}^{+} = 1$, and let $\mathcal{D}_i = \{j : \omega_{ij}^{+} > 0\}$. For support inputs $h_i \geq 0$, put $a_i^{+} = e_i - \omega_i^{+}$ and $\delta v = \sum_i h_i a_i^{+}$. Joint feasibility requires $\sum_i h_i \leq C^{+}$, $v + \delta v \geq 0$, and $h_i = 0$ outside the selected set and whenever $\widehat{\tau}_i \leq 0$. Define the true and score-based marginal net responses

$$\tau_i = \nabla \bar{U}_{t+1}(v)^{\top} a_i^{+} = \partial_i \bar{U}_{t+1}(v) - \sum_j \omega_{ij}^{+} \partial_j \bar{U}_{t+1}(v), \qquad \widehat{\tau}_i = \rho_i - \sum_j \omega_{ij}^{+} \rho_j. \tag{52}$$

Here $h_i \, \partial_i \bar{U}_{t+1}(v)$ is the first-order gross return, $h_i \sum_j \omega_{ij}^{+} \partial_j \bar{U}_{t+1}(v)$ its displacement cost, and $h_i \tau_i$ the first-order net gain. Eq. (38) replaces the marginal components by the value densities, so $\widehat{M}_i = h_i \widehat{\tau}_i$ and $\sum_i h_i \widehat{\tau}_i = \rho^{\top} \delta v$. If $\nabla \bar{U}_{t+1}$ is Λ-Lipschitz on the feasible segment and $|\widehat{\tau}_i - \tau_i| \leq \varepsilon_i$, then Eq. (40) holds. The proof first derives ε_i from the valuation pipeline and then controls the joint utility remainder. The valuation-based ε_i of Eq. (58) requires that every supported or displaced item entering the comparison has positive reference removal dose v_j equal to its density denominator c_j, and that the removal path satisfies the same smoothness condition. A zero denominator does not define a density score. For $v_j = 0$, a positive planned cost c_j defines a cold-start density P_j / c_j, but does not establish the shielding-based bound. The deployed prior gate does not invoke that bound.

From finite shielding to a local value density. For an item with positive reference exposure $v_j > 0$ and matched density denominator $c_j = v_j$, let ω_j^{-} describe where its exposure goes under shielding, with $\omega_j^{-} \geq 0$, $\omega_{jj}^{-} = 0$, and $\mathbf{1}^{\top} \omega_j^{-} = 1$. Put $a_j^{-} = e_j - \omega_j^{-}$ and define

$$V_j = \bar{U}_{t+1}(v) - \bar{U}_{t+1}(v - v_j a_j^{-}), \qquad \zeta_j = \sum_k \omega_{jk}^{-} \, \partial_k \bar{U}_{t+1}(v), \tag{53}$$

where V_j is the shielding effect of Eq. (1) at the common reference allocation and outcome window, and ζ_j is the replacement reference, the marginal value of the supply that absorbs the removed exposure. The denominator v_j is the reference removal dose of the value being compared, and the settled credit and the prior are expressed on the same utility and exposure scale. Under Λ-Lipschitz continuity of the gradient along the removal segment,

$$\frac{V_j}{v_j} = \frac{1}{v_j} \int_0^{v_j} \nabla \bar{U}_{t+1}(v - s a_j^{-})^{\top} a_j^{-} \, ds, \qquad \left| \frac{V_j}{v_j} - \big(\partial_j \bar{U}_{t+1}(v) - \zeta_j \big) \right| \leq \frac{\Lambda}{2} \, v_j \| a_j^{-} \|^2. \tag{54}$$

The compensation ratio of Section 2.2 is the model form of the replacement reference, $r_j = \zeta_j / \partial_j \bar{U}_{t+1}(v)$ for $\partial_j \bar{U}_{t+1}(v) \neq 0$, with the fitted marginal effects m in place of the marginal components of $\bar{U}_{t+1}$. When this denominator is zero, the net marginal response remains $\partial_j \bar{U}_{t+1}(v) - \zeta_j$ and is used directly without a ratio representation. Then $\partial_j \bar{U}_{t+1}(v) - \zeta_j = (1 - r_j) \, \partial_j \bar{U}_{t+1}(v)$, so the discounted gross term of Eq. (4) is the first-order net loss, and Eq. (54) bounds its distance from the shielding effect of a full removal. The identity holds for any real r_j. The restriction $r_j \in [0, 1]$ of Section 2.2 is the case in which ζ_j lies between 0 and $\partial_j \bar{U}_{t+1}(v)$, and it is exactly the case in which the net term keeps the sign of the gross term. Let R_j^{SCSV} be the settlement label and P_j the risk-adjusted reference score of Section 2.3. Define the per-exposure prediction and settlement discrepancies

$$\mathcal{E}_j^{\mathrm{prior}} = \frac{|P_j - R_j^{\mathrm{SCSV}}|}{v_j}, \qquad \mathcal{E}_j^{\mathrm{settle}} = \frac{|R_j^{\mathrm{SCSV}} - V_j|}{v_j}. \tag{55}$$

The first includes the effect of the risk adjustment, the second connects the operational credit to the shielding estimand, and neither is the uncertainty output of the prior model. Combining them with Eq. (54), the score density satisfies

$$\rho_j = \partial_j \bar{U}_{t+1}(v) - \zeta_j + \xi_j, \qquad |\xi_j| \leq \mathcal{E}_j^{\mathrm{val}} = \mathcal{E}_j^{\mathrm{prior}} + \mathcal{E}_j^{\mathrm{settle}} + \frac{\Lambda}{2} \, v_j \| a_j^{-} \|^2. \tag{56}$$

This separates prediction and settlement accuracy from the finite-removal approximation and uses the same ideal shielding quantity and settled credit as Sections 2.1 and 2.2, rather than assigning a new estimand to the prior.

Replacement references and boosting. Substituting Eq. (56) into Eq. (52) gives the identity

$$\widehat{\tau}_i - \tau_i = \xi_i - \sum_j \omega_{ij}^+ \xi_j - \zeta_i + \sum_j \omega_{ij}^+ \zeta_j, \tag{57}$$

and therefore

$$|\widehat{\tau}_i - \tau_i| \le \varepsilon_i, \qquad \varepsilon_i = \mathcal{E}_i^{\mathrm{val}} + \sum_j \omega_{ij}^+ \mathcal{E}_j^{\mathrm{val}} + \Big|\zeta_i - \sum_j \omega_{ij}^+ \zeta_j\Big|. \tag{58}$$

When the compared items share a common replacement reference, $\zeta_j = \zeta$, the last term vanishes. In general Eq. (58) measures the reference mismatch explicitly. This is the condition under which differences of substitution-corrected densities recover opportunity-cost comparisons. The replacement reference cancels or contributes a controlled residual, and is not counted as gross value.

Joint exposure increments. For jointly feasible increments $\delta v = \sum_i h_i a_i^+$, $\nabla \bar{U}_{t+1}(v)^\top \delta v = \sum_i h_i \tau_i$. Integrating along the joint segment,

$$\bar{U}_{t+1}(v + \delta v) - \bar{U}_{t+1}(v) = \sum_i h_i \tau_i + \mathcal{R}(\delta v), \qquad |\mathcal{R}(\delta v)| \le \frac{\Lambda}{2} \|\delta v\|^2, \tag{59}$$

and together with Eq. (58),

$$\Big|\bar{U}_{t+1}(v + \delta v) - \bar{U}_{t+1}(v) - \sum_i h_i \widehat{\tau}_i\Big| \le \sum_i h_i \varepsilon_i + \frac{\Lambda}{2} \Big\| \sum_i h_i a_i^+ \Big\|^2, \tag{60}$$

which is Eq. (40). □The joint remainder accounts for interactions among simultaneous increments. For a single item, a sufficient condition for positive expected gain is

$$\widehat{\tau}_i > \varepsilon_i + \frac{\Lambda}{2} h_i \|a_i^+\|^2. \tag{61}$$

Matched small interventions. If the same direction a_i is used for removing and adding a small dose h, write $R_i^-(h) = \bar{U}_{t+1}(v) - \bar{U}_{t+1}(v - h a_i)$ and $M_i^+(h) = \bar{U}_{t+1}(v + h a_i) - \bar{U}_{t+1}(v)$. The same argument gives $|M_i^+(h) - R_i^-(h)| \le \Lambda h^2 \|a_i\|^2$, which is the matched-direction reading of Appendix E. The density bridge applies to positive reference exposure. Very small positive exposure retains the derivation, although the prediction and settlement discrepancies divided by exposure may become large. For zero-exposure uploads, the prior transfers information from historical settlement labels, and the resulting intervention is assessed by the production experiments.

Operational interpretation. The exposure cap controls the size of the intervention, and the error bound additionally depends on the response curvature and on the accuracy and comparability of the value signal. Identifying the displacement weights specifies the opportunity-cost comparison, while Eq. (60) retains the remaining valuation and curvature errors. The inequalities give a conditional interpretation of the value-guided objective and do not add an error-bound estimator or a certified admission test to the deployed implementation.

G Causal Validity Checks

This appendix collects the checks supporting the causal reading of the settled credit. Appendix G.1 compares SCSV with conventional causal estimators on the same randomized ground truth and collects the bidirectional-consistency table of Section 3.2. Appendix G.2 reports placebo and falsification tests probing the identifying assumptions. Appendix G.3 reports correlational and structural checks, which are necessary but not sufficient for validity.

G.1 Comparison with conventional causal estimators

Common protocol. The log-based rows of Table 2 use the same observational logs, covariates $W_{a,t}$, temporal split, and shielded cohort, with the estimator-specific treatments and outcomes defined below. The RCT row is the randomized estimate itself, and the response projection row combines the same logged $\widehat{\theta}$ with experimental treatment contrasts. It is the window-level diagnostic of Appendix B.1 and is described under Response projection below.

Target. IPS and DR use the item-level share change $X_{a,i,t}$ as a scalar treatment and the linear marginal structural model

$$\mathbb{E}\big[Y_{a,t+1}(x)\big] = \vartheta_0 + \vartheta_1 x, \tag{62}$$

where $Y_{a,t+1}(x)$ is the next-period outcome when the share change of item i is set to x and the remaining exposure is reallocated according to the logging policy (Robins et al., 2000). The slope ϑ_1 is a single slope pooled over the items of the shielded cohort. It measures the net utility response per unit of exposure share, including the accompanying reallocation. For set shielding, an additive approximation applies this per-share response to every removed share and sums over the shielded items:

$$\widehat{L} = \widehat{\vartheta}_1\, b_{\mathcal{S}}, \qquad b_{\mathcal{S}} = \sum_a \sum_{i\in\mathcal{S}} b_{a,i,t}, \tag{63}$$

This prediction targets the total-scale counterpart of the cohort loss $L_{t+1}(z)$ in Eq. (14), provided the policy reallocates the removed share as the logging policy does, and under the same timing as Proposition B.2. The baselines use the change outcome $Y_{a,t+1}$ and the same $W_{a,t}$, so their slopes are also next-period responses at a fixed period-t level, and the same-period loss enters their comparison with the randomized window in the same way. The substitutes' terms, which SCSV carries in $\Phi_{a,t}$, are not modeled separately. The set prediction is linear and additive in the removed shares.

Treatment model. Because $X_{a,i,t}$ is continuous, the propensity-based estimators use a generalized propensity score (Hirano & Imbens, 2004). The conditional density $e(x \mid W_{a,t}) = p(X_{a,i,t} = x \mid W_{a,t})$ is estimated with a Gaussian working model whose mean is a regression on $W_{a,t}$ and whose variance is the residual variance, and the marginal density $e(x)$ with the same working model without covariates. The fit uses the pairs (a, i, t) with i in the shielded cohort and $i \in \mathcal{C}_{a,t}$, weighted by

$$w_{a,i,t} = \frac{\widehat{e}(X_{a,i,t})}{\widehat{e}(X_{a,i,t} \mid W_{a,t})}, \tag{64}$$

the inverse conditional density stabilized by the marginal density (Robins et al., 2000).

Estimators. IPS fits Eq. (62) by weighted least squares,

$$\big(\widehat{\vartheta}_0^{\mathrm{IPS}}, \widehat{\vartheta}_1^{\mathrm{IPS}}\big) = \arg\min_{\vartheta_0,\vartheta_1} \sum_{(a,i,t)} w_{a,i,t}\big(Y_{a,t+1} - \vartheta_0 - \vartheta_1 X_{a,i,t}\big)^2, \tag{65}$$

and is consistent for ϑ_1 when the density model is correctly specified (Robins et al., 2000). The propensity-weighted value estimator solves Eq. (65) with the same weights and with $Y_{a,t+1}$ replaced by the engagement attributed to item i in unit a in period $t+1$, in the same utility metric. Its slope is the attributed engagement per unit of share, the closest propensity-based analogue of an engagement-attributed value, and its $\widehat{L}$ is the attributed engagement removed with the share. DR fits an outcome model $\widehat{\varrho}(x, w)$ for $\mathbb{E}[Y_{a,t+1} \mid X_{a,i,t} = x, W_{a,t} = w]$ directly on $Y_{a,t+1}$ and regresses on $X_{a,i,t}$ the pseudo-outcome

$$Y^{\dagger}_{a,i,t} = w_{a,i,t}\big\{Y_{a,t+1} - \widehat{\varrho}(X_{a,i,t}, W_{a,t})\big\} + \overline{\widehat{\varrho}}(X_{a,i,t}), \qquad \overline{\widehat{\varrho}}(x) = \frac{1}{n}\sum_{(a',t')} \widehat{\varrho}(x, W_{a',t'}), \tag{66}$$

in which the cross-fitted $\widehat{\varrho}$ is held fixed and $\overline{\widehat{\varrho}}$ averages it over the n sampled covariate vectors. The first term is the inverse-density-weighted residual correction, the second the outcome model. $\mathbb{E}[Y^{\dagger}_{a,i,t} \mid X_{a,i,t} = x] = \mathbb{E}[Y_{a,t+1}(x)]$ if either the density model or the outcome model is correctly specified (Kennedy et al., 2017), so $\widehat{\vartheta}_1^{\mathrm{DR}}$ is consistent under either. SCSV is the orthogonalized estimator of Appendix B followed by the settlement of Eq. (6). All nuisance functions, including the density model, the outcome model, and the SCSV nuisance models $\widehat{m}(\cdot)$ and $\widehat{\ell}(\cdot)$, use gradient-boosted trees with $K = 5$ cross-fitting folds, and all estimators share the same fold assignment, so the comparison does not favor SCSV through a richer nuisance fit.

Bidirectional consistency. Table 10 collects the online effects and offline estimates of the two intervention directions compared in Section 3.2.

Table 10: Intervention consistency at the author-cohort aggregation level. Effects are treated-minus-control relative changes.

Intervention	Online	Offline estimate
Boosting	$+0.230\%$	$+0.21\%$
Shielding	-1.351%	-1.350%

Aggregation to the shielded set. Each baseline scales its fitted slope by the removed share $b_{\mathcal{S}}$ (Eq. (63)). SCSV sums the settled credits of the shielded items (Appendix J). Each resulting prediction is negated and divided by the settlement-period open-reason total U_{ref}, yielding the shield-minus-control relative changes of Table 2 (Eq. (72)). The randomized row is the platform's difference-in-differences relative lift over the experiment window, normalized by the contemporaneous control arm, an estimate of Δ_{win} in Eq. (19). The projection row is normalized by the same control-arm mean (Eq. (68)). All rows share the utility metric and the full-page measurement scope. The log-based rows are relative losses at the settlement period and the two experiment-based rows are relative losses over the window. The comparison between the two groups rests on condition (f′) of Appendix B.1, and no row is rescaled to another row's utility level.

Uncertainty. Standard errors are computed by a cluster bootstrap over attribution units with $B = 1000$ resamples. The projection row clusters on author buckets, whereas the other rows cluster on attribution units. The relative variances implied by Table 2 are 9.0 for IPS, 7.4 for the propensity-weighted value estimator, and 2.25 for DR, with SCSV normalized to 1.

Response projection. The last row of Table 2 is a window-level diagnostic motivated by Eq. (16). $\widehat{\theta}$ is the estimate of Eq. (3) above, and $\widehat{d}_t$ is the difference between the control and shielded arms in the per-unit mean of $\Phi_{a,t}$, built with the same $\phi_{a,i,t}$, period definition, and unit as the log fit and weighted by the experimental design. No gate, discount, or budget is applied. The log fit uses daily periods. For the experiment we define a window contrast with t the first shielded period and $t-1$ the last pre-shielding period, as in Appendix B.1. The contrast uses $X^{\mathrm{win}}_{a,i,s} = b_{a,i,s} - b_{a,i,t-1}$ for each period s of the window, $\Phi^{\mathrm{win}}_{a,s} = \sum_{i\in\mathcal{C}}(s_{a,i,s} - s_{a,i,t-1})$, whose share reading is $X^{\mathrm{win}}_{a,i,s}\phi_{a,i,s}$ as in Section 2.2, and $\widehat{d}^{\mathrm{win}}_s$, the between-arm difference in the mean of $\Phi^{\mathrm{win}}_{a,s}$. In the first period $X^{\mathrm{win}}_{a,i,t} = X_{a,i,t}$ and $\widehat{d}^{\mathrm{win}}_t = \widehat{d}_t$, the contrast of Proposition B.2. In later periods Eq. (2) records no further change for a removed item, whereas X^{win} keeps the change that the fixed policy maintains. The projection is $\widehat{\theta}^{\top}\overline{\widehat{d}}^{\mathrm{win}}$, where $\overline{\widehat{d}}^{\mathrm{win}} = H^{-1}\sum_{s=t}^{t+H-1}\widehat{d}^{\mathrm{win}}_s$ averages the treatment contrast over the H shielded days. In the open-reason experiment $H = 5$, preceded by a 7-day pre-period, and the platform's difference-in-differences relative lift is taken over the same 5 shielded days. Under the lag convention these treatment days predict the outcomes of $t+1,\dots,t+H$, one period later than the window of Eq. (19). The offset matters at the two edge periods. The platform's window contains the first shielded day, whose loss is the same-period term $L_t(z)$ of Proposition B.2 rather than a lagged response, and excludes period $t+H$, on which the lagged response of the last shielded day falls. Both edges are visible only in the daily series. The weight of the first under (f) is given by Eq. (20), and in the window average their contribution is part of the joint diagnostic below. The decomposition of the projection is

$$\widehat{\theta}^{\top}\overline{\widehat{d}}^{\mathrm{win}} = \widehat{\theta}^{\top}\widehat{d}_t + \widehat{\theta}^{\top}\big(\overline{\widehat{d}}^{\mathrm{win}} - \widehat{d}_t\big). \tag{67}$$

The projection is a per-unit, per-day loss over the window, and its signed relative form is

$$\widehat{\Delta}^{\mathrm{win}}_{P,\mathrm{shield}} = -\frac{\widehat{\theta}^{\top}\overline{\widehat{d}}^{\mathrm{win}}}{\bar{\mu}^{0}_{\mathrm{win}}}, \tag{68}$$

where $\bar{\mu}^{0}_{\mathrm{win}} > 0$ is the control arm's per-unit, per-day utility averaged over the H shielded days, the denominator of Δ_{win} in Eq. (19), for the same metric and measurement scope. A mean is divided by a mean. A cumulative H-day numerator would instead require the H-day control total. The reported

projection is -1.300% with SE 0.16%, 0.051 pp from the randomized effect and 0.05 pp from the signed SCSV estimate in the row above.

In Eq. (67), the first term is the full first-period projection $\widehat{\theta}^\top \widehat{d}_t$, the lagged-response term of Proposition B.2, whose component along ι is identified only under (b_0) (condition (e)). The second term is the drift of the treatment contrast over the window. Removed items stay at zero exposure in the treated arm, but the control arm keeps moving, substitutes respond, and $\phi_{a,i,s}$ changes from day to day, so the between-arm contrast is not constant even under a fixed policy. Condition (f) concerns outcome losses normalized by contemporaneous control utility and says nothing about the treatment contrast. A flat daily arm gap, even where observed, would not establish (f). The projection and the randomized estimate $\widehat{\Delta}_{\mathrm{RCT}}$ of Table 2, which estimates Δ_{win} of Eq. (19), share the metric scope, the window, and the control-arm denominator $\bar{\mu}^0_{\mathrm{win}}$, and their signed difference $\widehat{\Delta}^{\mathrm{win}}_{P,\mathrm{shield}} - \widehat{\Delta}_{\mathrm{RCT}}$ is $(-1.300) - (-1.351) = +0.051$ pp (SE 0.140 pp). This is a joint diagnostic of the first-period bridge, the same-period loss of the first day, drift of the treatment contrast, and changes in the relative outcome loss within the window. A small total difference may reflect cancellation among these terms. Standard errors are standard deviations over $B = 1000$ cluster bootstrap replicates. In each replicate the same author-bucket draw is applied to the log data and to the experiment data. Each replicate re-estimates $\widehat{\theta}$, recomputes the arm-specific window contrasts, and jointly recomputes the normalized projection and the randomized estimate under the same reference convention. The standard error of their signed difference is computed directly from the replicate-wise differences, preserving the dependence between the two estimates. This gives 0.160 pp for the projection and 0.140 pp for the difference, with an approximate 95% interval of $[-0.223,\ 0.325]$ pp for the difference. Because $\widehat{d}_t$ comes from the experimental arms, the standard error of the projection contains their sampling variation and is not comparable one to one with the log-only bootstrap of the other rows.

Same-period and lagged components. Proposition B.2 writes the next-period loss as the same-period loss $L_t(z)$ carried into the level, the lagged response $\theta^\top d_t(z)$, the level term $\Delta f_t(z)$, and the residual η_t, which under a policy kept in force contains the same-period loss of the second shielded day. The window comparison above does not separate them. The window opens on the first shielded day, whose loss is same-period by construction, and the projection uses the full contrast, whose component along ι is that same-period loss. The daily series of an experiment gives $L_t(z) = \iota^\top d_t(z)$ directly, since the first shielded day carries only the same-period loss, and gives $L_{t+1}(z)$ as the arm difference of the second day. Splitting $L_{t+1}(z)$ into its lagged, level, and same-period parts requires the model, and the comparisons reported here use the window average. The two validation metrics differ in how much of a sustained loss is next-period. An open reason is a user-initiated entry attributed to a revisit, share, follow-page, or search session (Section 2.1), an action that follows exposure in the feed, so a larger part of its response to a removal from the feed is lagged than for watch time, which responds within the period. The agreement of the settled estimate with the randomized effect on open reasons in Table 2, and the shortfall of the settled estimates below the online watch-time losses at every ratio in Table 1, are both consistent with this ordering. Neither is a measurement of $L_t(z)$, which only the daily series gives.

Scope. The evaluation compares set-level predictions with randomized shielding effects. All three baseline predictions lie within one of their respective bootstrap standard errors of the randomized estimate, with standard errors 1.5 to 3 times that of SCSV. SCSV also assigns item-level credit subject to the unit-level settlement budgets.

G.2 Placebo and falsification tests

We probe the identifying assumptions of Section 2.2 and Section 4 directly.

Reversed timing. The attribution design regresses the outcome change over $(t, t+1)$ on treatment changes over $(t-1, t)$. We re-run Eq. (3) with the roles reversed, using future treatment changes as regressors for the preceding outcome change, with identical covariates, folds, and regularization. The reversed regression yields $+0.003\%$ (SE 0.019%, $p = 0.87$). A same-period feedback loop between allocation and utility would tend to produce a nonzero reversed association. None is detected at the reported precision.

Table 11: Segment-level alignment within 7 days post posting. Across high-value segments, we compare volume share, settled-credit share, and online open-reason share.

Items Segment	Volume	Settled credit	Online
50k breakout	0.12%	32.44%	21.46%
UE score ≥ 6	2.17%	43.74%	37.42%
10k+ fans author	9.98%	59.77%	56.58%

Pre-period trends. The DiD estimates in Table 4 require that the arms would have evolved in parallel absent intervention. Over the 14-day pre-period, the tests detect no differential trends in view time ($+0.004\%$, $p = 0.79$) or retention ($+0.002\%$, $p = 0.88$).

Permutation. Reassigning treatment labels across randomization units 1,000 times and recomputing the DiD statistic yields a null distribution centered at -0.001% (95% range $[-0.037\%, +0.036\%]$). The observed view-time effect of $+0.053\%$ falls outside this range (empirical $p < 0.01$).

Negative controls. Client crash rate and network latency, which the ramp-up mechanism is not expected to affect, show no association with treatment (-0.001%, $p = 0.94$).

None of these checks rules out unobserved confounding. The timing, pre-period, and negative-control checks detect no statistically significant departures from their null hypotheses at the reported precision, and the permutation test places the observed view-time effect outside the central 95% of the null distribution. The randomized comparisons of Section 3.1 and Section 3.2 provide separate evidence. The offline predictions are close to the observed shielding loss (-1.350% versus -1.351%) and promotion gain ($+0.21\%$ versus $+0.230\%$).

G.3 Correlation and structural sanity checks

Finally, we complement the causal validation with auxiliary analyses assessing the signal fidelity and structural consistency of the settled credit. While we re-emphasize that observed engagement proxies are not causal ground truth under strong compensation effects, mid-to-high correlation remains a necessary (though not sufficient) condition for signal validity.

Signal fidelity. We assess the fidelity of the posterior value score against observed "content open reasons". The analysis reveals strong alignment, with overall Spearman rank correlations reaching 72.90% and Pearson value correlations up to 82.32% across key dimensions. This indicates that the settled credit reliably ranks content relative to observed engagement proxies, despite the theoretical limitations of proxy-based validation.

Structural consistency. We further examine whether the distribution of estimated value aligns with delivery structures and realized engagement across strategic segments within 7 days of posting. As presented in Table 11, the settled credit captures the highly skewed nature of value concentration across diverse segment definitions. For instance, the `50k breakout` segment, representing a mere 0.12% of content volume, is identified as contributing 32.44% of total value, a structure that broadly matches its 21.46% share of realized 7-day open reasons. Similar structural alignment is observed across high-quality item segments (content-quality (UE) score ≥ 6) and top-tier creator cohorts (e.g., 10k+ fans authors), validating that the settled credit correctly localizes high-value pockets in agreement with empirical observation. This segment-level concentration shows how settled credit departs from volume. A small-volume, high-leverage segment can carry a disproportionate fraction of the settled credit, which is exactly the regime where engagement-based attribution fails.

H Model and Training Details

Model and evaluation setup. We instantiate the posterior value pipeline with a linear DML design. Inputs include a treatment tower (historical content-side value signals such as sharing and revisits), backdoor covariates (user-side attributes such as age), and a target defined on future content value.

Training uses a staged fit with "backdoor-first" residualization before fitting the treatment component. This staged fit is the production counterpart of the residualization of Appendix B. The nuisance towers are fit first and held fixed, as Fig. 2 indicates, while the linear attribution layer is fit on their residuals. The estimator comparison of Appendix G.1 replaces the towers with gradient-boosted trees for every method, so that comparison is not driven by nuisance capacity. Under causal-stability constraints, the reported Kendall rank agreement is 0.63, indicating non-trivial rank fidelity for downstream selection and aggregation use.

Dataset construction and split definition. We build an anonymized, day-indexed observational dataset from production logs through a deterministic four-stage SQL pipeline, organized in three logical layers for feature aggregation, posterior label construction, and temporal splitting.

Feature aggregation. The atomic record is defined at the *user–item–day* level, indexed by user ID, item ID, and date, and stores impression counts together with eight score-weighted behavior signals, including play duration, likes, follows, shares, comments, open-reason session attribution, reduce-similar actions, and short plays. The open-reason session feature aggregates revisit, share, follow-page, and search session attribution. These records are then rolled up to the *user–day* level to form user-side covariates, such as activity-level buckets, and are further joined with fixed item-, author-, and photo-level feature sources.

Posterior label construction. Labels are produced over a three-day temporal window, d_1, d_2, d_3, which are the periods $t-1$, t, $t+1$ of Section 2.2 (Table 9). Behavior deltas are computed from d_1 to d_2, and the future outcome delta $Y_{a,t+1}$ is computed from d_2 to d_3. The covariates $W_{a,t}$ combine the lookback features assembled before d_2 with the utility level of d_2 (Appendix B). In this implementation the treatment term of Section 2.2 is the change in the item's value-weighted behavior vector between d_1 and d_2. An item consumed on d_2 and absent on d_1 contributes its d_2 vector, and an item consumed on d_1 and absent on d_2 contributes the negative of its d_1 vector. An item consumed on both days contributes the change in its vector, which carries its share change at its d_2 profile together with the change in its per-share behavior at its d_1 share. This lagged design decouples treatment-side variation from its subsequent response and mitigates contemporaneous exposure leakage. The eight behavior signals are converted into a unified behavior-to-value score using fixed production-calibrated weights. The normalized user-level outcome delta is then attributed back to individual items through the unit-level rule of Section 2.2. In this alpha-based sign-consistent rule, only item-level score changes aligned with the direction of the user-level change receive credit (the gate), and each admitted mass is discounted by $1 - r_{i,t}$ (Appendix J). The credit is then scaled by the capped attribution coefficient $\alpha_{a,t}^{\mathrm{cap}}$ and oriented by the reference score $\sigma_{a,i,t}$ of Eq. (7).

Temporal split and evaluation. Examples are partitioned by date and explicitly tagged as `train` or `eval`. Features are assembled with fixed temporal offsets, primarily a seven-day lookback from the partition date, while labels are derived from delayed posterior attribution. For evaluation, we use randomized supply-shielding ground truth. Selected items are randomly removed from recommendation in online experiments, and the resulting engagement change is compared with the offline posterior-value estimate to test whether the offline estimate recovers the set-level shielding loss under substitution compensation.

Hyperparameters and training setup. Table 12 summarizes the main data-processing, model, optimization, and training settings used in the posterior-to-prior modeling pipeline.

Hardware and compute. The data construction pipeline is executed on a Spark cluster with 100 executors, each using 8 cores, and takes approximately 4-8 hours per daily run, corresponding to about 4,800 core-hours per day. Model training is CPU-based. A full cold-start training run requires approximately 2 CPUs for 48 hours, plus validation, resulting in about 112 CPU-hours per full run. Daily incremental training requires approximately 12 CPU-hours, including both training and validation. The online inference path is lightweight, with measured per-request latency of 20 ms at P50 and 78 ms at P99.

Anonymized core pseudocode. To make the implementation logic reproducible while preserving production anonymity, we provide anonymized pseudocode for the core estimation and settlement

Table 12: Main hyperparameters and training setup.

Component	Setting
Spark pipeline	Driver memory 10GB (3GB overhead); executor memory 20GB (10GB overhead); 8 executor cores; max input split size 4GB.
Value-score construction	Eight fixed behavior-to-value weights are used for play duration, likes, follows, shares, comments, open-reason sessions, reduce-similar actions, and short plays, with normalization constant 0.3353.
Input representation	Approximately 380-dimensional input, combining numeric features, categorical embeddings, pretrained content embeddings, compressed quality embeddings, and context-attention features.
Model backbone	Multi-gate Mixture-of-Experts (MMoE) with 6 experts and 3 task gates. Each expert uses hidden sizes $[128, 64]$ with ReLU activation and softmax gating.
Context attention	Transformer-style attention with 4 heads and 16-dimensional attention embeddings over contextual supply sets, using 50 context items and 200 similar-tier items.
Value heads	Three value-prediction heads: Value Final, PC Value Final, and OR Value Final. Each uses a 64-dimensional task tower with sigmoid prediction and softplus evidence outputs.
Evidential regression	$\lambda_{\text{ev}} = 0.01$, $\epsilon = 10^{-8}$, squared-error clipping at 100, and loss clipping to $[-50, 50]$.
Optimization	Sparse and dense parameters are optimized separately with Adam, both using learning rate 10^{-3}.
Loss weights	DER loss weights are 1.0, 1.3, and 1.0 for Value Final, PC Value Final, and OR Value Final, respectively.
Training workflow	SQL data construction, feature-reader generation, model training, evaluation, and serving export.

pipeline. The pseudocode abstracts away platform-specific identifiers, storage paths, and serving infrastructure, but keeps the algorithmic components required to reproduce the method. It includes four modules. The first is a value-attribution pipeline that computes behavior scores, constructs the behavior-vector treatment terms and the three-day outcome delta, selects sign-consistent item changes, applies the replaceability discount, the capped attribution coefficient, and the reference-score orientation of Eq. (7), and aggregates user-level attributions into item-level posterior values. The second is a Deep Evidential Regression module that computes evidence-based uncertainty and the uncertainty-aware loss. The third is an MMoE-based prior scorer that encodes item, context, and quality features and outputs value predictions with evidence. The fourth consists of training and inference routines that specify separate sparse/dense optimization, metric computation, and serving-time uncertainty-aware prediction. This pseudocode is intended to document the core estimation and settlement logic. The complete core pseudocode is provided in the supplemental material (*pseudocode_core_pipeline.py*).

I Additional Prior-Scoring and Compensation Diagnostics

This appendix provides additional diagnostics for the serving-time prior scorer used in Section 3.3. The main text focuses on intervention-readiness, namely rank preservation, budget leverage, and reduced compensation. Here we report the supporting distributional diagnostics and compensation measurements.

Scope of the prior-scoring evaluation. Sections 3.1 and 3.2 validate the ex-post SCSV settlement against interventional ground truth, and Section 3.3 evaluates whether the substitution-corrected posterior value can be distilled into a serving-time prior score that is suitable for value-guided intervention under finite attention. This appendix reports the diagnostics behind that evaluation. Following the intervention formulation in Section 2.3, we focus on three properties required by an intervention-ready value signal: (i) rank preservation under production shift, (ii) budget leverage under monotone gating, and (iii) reduced compensation when the score is used for supply selection. Uncertainty is used only as a conservative serving control. Multi-day distributional diagnostics follow below.

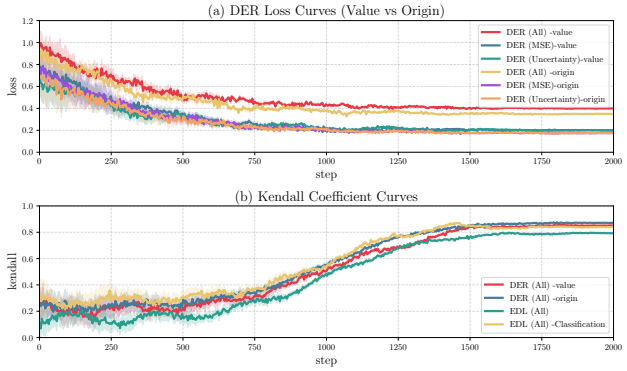


Figure 5: Posterior-to-prior value distillation.

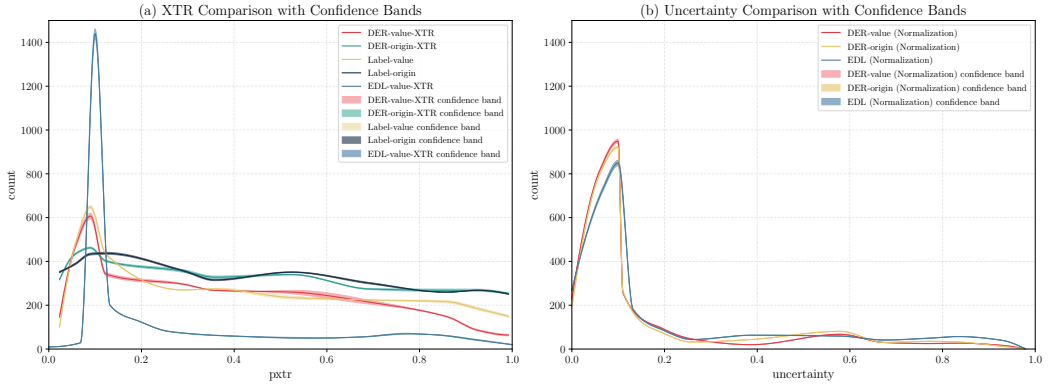


Figure 6: Multi-day stability of the prior value score and uncertainty under production shift. The score distribution remains stable across production days, while the uncertainty signal provides a conservative control surface for cold-start ramp-up.

Rank fidelity for serving-time selection. Fig. 5 examines whether the prior model preserves the serving-relevant ordering of the compensated SCSV label. Panel (a) shows that deep evidential regression (DER) can stably optimize both the compensated value label and the uncompensated origin label, indicating that compensation-aware supervision does not introduce optimization instability. Panel (b) provides the ranking evidence. Using Kendall's τ against the continuous compensated-value label, DER (Amini et al., 2020) reaches $\tau = 0.852$, compared with $\tau = 0.793$ for evidential deep learning (EDL) (Sensoy et al., 2018). EDL is trained on a thresholded version of the same label (see the model-choice diagnostics below) and evaluated on the same ranking criterion. This distinction matters because the online intervention policy uses the prior through an uncertainty-aware eligibility gate and a value-based boost feature, rather than through binary classification accuracy. Fig. 5 supports DER as a stable implementation of posterior-to-prior distillation. What it measures is rank fidelity to the settled label. Items whose labels differ by more than the sum of their prediction errors on the training scale keep their order under the prior. Whether that order tracks shielding value is a property of the label (Section 3), and the value of the supply selected by the prior is reported in Table 3.

Multi-day stability under production shift. Because serving consumes (π_i, u_i) through monotone thresholds and ranking, we evaluate not only pointwise fit but also distributional stability. For each production day, we compute the empirical distribution of the normalized prior score and uncertainty, then aggregate across days by plotting the mean curve with a 95% confidence band. This protocol probes distribution shift by checking whether the score distribution remains well-shaped across days and whether uncertainty mass inflates in the long tail.

Fig. 6 summarizes the multi-day distributions. Compared with EDL trained with a thresholded classification-style target, DER produces a score distribution closer to the continuous compensated-

value label and avoids excessive mass collapse near 0 or 1, which is critical for threshold-based gating. The uncertainty distributions remain aligned and stable across days, supporting the use of u_i as an auditable risk-control surface. In contrast, EDL's uncertainty becomes less discriminative under the joint effect of thresholding and auxiliary objectives, making it harder to separate genuinely uncertain long-tail items from in-distribution head items.

Additional model-choice diagnostics. The full training diagnostics show that lower classification loss does not necessarily imply better serving-time value ranking. EDL can obtain lower cross-entropy loss because the binary target is easier, but this does not resolve the score-shape mismatch induced by thresholding. The decisive metric for value-guided intervention is Kendall rank agreement with the continuous compensated-value label, since the online policy consumes the prior through an uncertainty-aware eligibility gate and a value-based boost feature. This is why Section 3.3 adopts DER as the default implementation of posterior-to-prior distillation.

Budget leverage under monotone gating. The first requirement of positive supply intervention is to identify a compact candidate set under a finite attention budget. We therefore analyze the serving behavior of π_i under the monotone eligibility gate used by the intervention policy. A threshold sweep reveals strong value concentration. Admitting only 7.49% of items covers 29.07% of the combined content-plus-friends open-reason endpoints, corresponding to approximately $3.9\times$ concentration. Similar amplification appears across decomposed endpoints. This indicates that the prior score transforms the posterior SCSV settlement into a high-leverage candidate pool, supporting score-based candidate prioritization under the ramp-up budget of Section 2.3.

J Offline shielding ablation and compensation diagnostics

We compare an uncompensated baseline score with the compensation-aware score via an offline shielding ablation. For each score, we construct a shielded set by selecting the highest-leverage supply under that score and measure the induced system-level change on open-reason endpoints under full-page scope shielding.

Set-level estimate. For a shielded set $\mathcal{S}$, the settlement-based estimate on the control-minus-shield loss scale is

$$\widehat{L}(\mathcal{S}) = \sum_{i\in\mathcal{S}} R_{i,t}^{\mathrm{SCSV}}(\mathcal{S}), \qquad R_{i,t}^{\mathrm{SCSV}}(\mathcal{S}) = \sum_{a} R_{a,i,t}(\mathcal{S}), \tag{69}$$

where $R_{a,i,t}(\mathcal{S})$ is the unit-level settlement of Eq. (5) computed with the set-specific masses $q_{a,i,t}(\mathcal{S}) = (1 - r_{i,t}(\mathcal{S}))\,|c_{a,i,t}|\,\mathbf{1}\{c_{a,i,t}Y_{a,t+1} > 0\}$ from production logs without use of the experiment's outcomes, and U_{ref} is the total of the same metric over the attribution units and settlement period of the logs from which $\widehat{L}(\mathcal{S})$ is settled (Appendix G.1). When the settlement window spans several periods, $\widehat{L}(\mathcal{S})$ and U_{ref} are summed over the same periods. The set enters through membership and through the compensation ratios $r_{i,t}(\mathcal{S})$, which are evaluated on the compensation base of Section 2.2 that remains after $\mathcal{S}$ is removed (Eq. (70)). The attribution weights, the sign gate, and the unit budgets $|Y_{a,t+1}|$ are the same for every evaluated set. The estimate is expressed in the utility units of the logs and is bounded by the realized unit-level changes, $|\widehat{L}(\mathcal{S})| \le \sum_a |Y_{a,t+1}|$ (Proposition 2.1). The calibration of $r_{i,t}$, described below, uses shielding rounds that precede and are disjoint from the experiments of Section 3, and no other parameter of the pipeline is tuned on those experiments. Since nothing else is tuned on them, and no scale parameter is fitted, an offline estimate is a settlement-based prediction rather than a fit, and the deviations reported in Section 3 measure its error against independent randomized effects. Appendix B.1 gives the cohort-level counterpart $\theta^\top d_t(z) = A_\theta(1 - r_\theta)$ of the lagged projection, in which the same dependence appears through r_θ.

Compensation-ratio estimation. The compensation ratio is a parametric function of the compensation base. Let $\mathbf{d}_{i,t}^{\mathrm{sub}}(\mathcal{S})$ collect substitute-density features of item i computed on the compensation base of Section 2.2. In the production settlement, that base consists of the items of the unit that are not credited. Under a shielding policy, it is the supply that remains after $\mathcal{S}$ is removed. The features measure how much of that base lies in the content neighborhood and creator cohort of item i, in exposure share and in item count. The ratio is

$$r_{i,t}(\mathcal{S}) = \min\big\{1, \max\{0, \varsigma^\top \mathbf{d}_{i,t}^{\mathrm{sub}}(\mathcal{S})\}\big\}, \tag{70}$$

so the restriction $r_{i,t} \in [0,1]$ of Section 2.2 is imposed by the clipping. One parameter vector ς is shared across items, periods, and utility metrics. It is calibrated on $N_{\mathrm{cal}} = 10$ randomized shielding rounds that precede and are disjoint from the experiments of Section 3. The number of rounds exceeds the dimension of ς, so the fit below is not underdetermined. Round k shields a set $\mathcal{S}_k$ and reports the platform's difference-in-differences relative lift $\widehat{\Delta}_k$, so its observed relative net loss is $\lambda_k^{\mathrm{obs}} = -\widehat{\Delta}_k$. Its settlement-based counterpart is the offline estimate of Eq. (69), settled with $r_{i,t}(\mathcal{S}_k)$ through the gate and the unit budgets and divided by the reference total of the round's logs, $\widehat{\lambda}_k(\varsigma) = \widehat{L}(\mathcal{S}_k;\varsigma)/U_{\mathrm{ref},k}$. The calibration solves

$$\widehat{\varsigma} \in \arg\min_{\varsigma} \sum_{k=1}^{N_{\mathrm{cal}}} \big(\widehat{\lambda}_k(\varsigma) - \lambda_k^{\mathrm{obs}}\big)^2, \tag{71}$$

and $\widehat{\varsigma}$ is frozen before the experiments of Section 3. The target is the observed net loss, which under conditions (b_0) and (c) of Appendix B.1 has the projection form $A_\theta - H_\theta$ of Eq. (17) and otherwise also contains the same-period and level terms of Eq. (16). A shielding round identifies only the net loss, and the gross term and the recovered compensation are not matched separately. Because the objective is evaluated after the gate and the budget, ς is informed by units whose admitted mass falls short of their realized change and by the shares of admitted items within a unit. In a unit whose budget binds and whose admitted items share one ratio, the ratio cancels. Under condition (f′) of Appendix B.1, $\widehat{\lambda}_k(\varsigma)$ and λ_k^{obs} target the same relative loss. The supervision is at the level of sets, so $\widehat{\varsigma}$ is identified through set-level agreement and no item-level ratio is validated individually (Section 2.1).

Locality of the calibrated ratio. Within the response model the identity between the net contribution and the gross contribution times $1 - r_{i,t}$ (Section 2.2) holds for any size of exposure change, because the model is linear in $\Phi_{a,t}$. The ratio varies with a removal only through where the withdrawn share goes, and the model's own extrapolation to a full removal is the locality of θ noted in Section 2.2, bounded in Appendix F.2. Since the calibration rounds remove whole sets, the calibrated ratio is the model ratio at that dose together with whatever the linear response misses there. Its use at the budget-capped changes of the logs rests on the locality of the response in the direction opposite to the extrapolation of θ, and is checked only at the level of sets (Section 3).

Reporting convention. The signed loss estimate $\widehat{L}(\mathcal{S})$ is positive when the settlement predicts that shielding reduces utility. Tables 1 and 2, and the shielding row of Table 10, report the opposite orientation:

$$\widehat{\Delta}_{\mathrm{shield}}(\mathcal{S}) = -\frac{\widehat{L}(\mathcal{S})}{U_{\mathrm{ref}}}, \qquad U_{\mathrm{ref}} > 0. \tag{72}$$

The ratio is dimensionless and displayed as a percentage. A positive loss produces a negative entry. Online effects are the platform's difference-in-differences relative lifts. Each lift is the treated-minus-control difference in per-unit utility over the experiment window, divided by the contemporaneous control-arm utility and net of the pre-period arm gap (Appendix L). These lifts estimate Δ_{win} of Eq. (19). The full-page measurement scope of the open-reason experiments covers the major traffic surfaces (Friends tab, social page, follow page, local feed, and single-/double-column feeds). Each relative change is thus normalized within its own data source and period. An offline estimate is compared with an online effect as a predicted against a realized relative loss under condition (f′) of Appendix B.1, with no rescaling between the two utility levels (Eq. (21)). This sign conversion does not assert that settled credit identifies the causal loss. The boosting row reports its own treated-minus-control gain, constructed in Section 3.2. It is not obtained by negating the shielding estimate. Absolute deviations and standard errors are unchanged by reversing signs, while signed differences reverse when converted between change and loss scales.

The numerator and reference total in Eq. (72) must use the same utility metric, measurement scope, period, and time aggregation. For a single period covering N_{ref} equally weighted attribution units, $\bar{u}_{\mathrm{ref}} = U_{\mathrm{ref}}/N_{\mathrm{ref}}$ gives $-\widehat{L}/U_{\mathrm{ref}} = -(\widehat{L}/N_{\mathrm{ref}})/\bar{u}_{\mathrm{ref}}$, so totals are compared with totals, or means with the corresponding means. Appendix G.1 gives the per-unit projection and its window normalization.

Shielding efficiency. We run an offline shielding ablation to test whether the compensation-aware score preferentially selects low-replaceable supply under substitution (Section 3.3). Compensation-

aware selection is more efficient and concentrates less replaceable supply. Shielding 13% of supply chosen by the compensation-aware score yields a 2.167% drop on the combined content+friends open reason, while shielding 22% chosen by the uncompensated baseline yields only a 1.924% drop. The SCSV-selected set therefore produces a larger loss per unit of removed supply at the evaluated selection rates.

B-to-C discount coefficient. To quantify compensation explicitly, we report a B-to-C discount coefficient, the ratio of supply-side (B) attributed mass to the consumer-side (C) system-level change:

$$\texttt{B2C}(e) \triangleq \frac{\text{AttributedShare}_B(\mathcal{S}_{\text{shield}};\, e)}{|\Delta_{\text{system}}(e; \mathcal{S}_{\text{shield}})|}, \tag{73}$$

where $\text{AttributedShare}_B(\mathcal{S}_{\text{shield}};\, e)$ is the B-side attributed mass of endpoint e captured by the shielded set $\mathcal{S}_{\text{shield}}$, and $\Delta_{\text{system}}(e; \mathcal{S}_{\text{shield}})$ is the system-level change induced by shielding that set. Smaller values indicate weaker compensation (lower replaceability), because the same attributed mass translates into a larger system-level impact when removed supply is less substitutable.

Across key endpoints, compensation-aware selection consistently yields smaller coefficients than baselines. For content open reasons, the coefficient decreases from 100.05 under a random baseline to 38.477 under the value-guided selection. These set-level comparisons associate value-guided selection with a smaller gap between attributed value and system-level shielding loss.

Dose profile of the watch-time series. This paragraph expands the discussion of Table 1 in Section 3.1. In the model, the estimate for ratio p depends on p only through which items enter the shielded set and through $r_{i,t}$, which is evaluated on the compensation base that remains after shielding (Appendix J). With $r_{i,t}$ held fixed the estimate would be linear in p under random membership, so the offline curvature is produced by the compensation ratio falling as the shielded set grows. Small removals are absorbed by substitutes, whereas larger removals deplete them. The online profile is consistent with this account. The 90% loss is 8.9% on the main app and 8.8% on lite. The absolute deviation lies between 0.27 and 0.67 percentage points, and the relative deviation falls from 77% at 10% shielding to about 6% at 90%. Every offline estimate is smaller in magnitude than the corresponding online loss. The gate, the discount, and the budget only remove credit, so the estimate cannot exceed the gross attributed mass. In this series it also falls short of the realized loss, most at the low end where the shielded supply is the most replaceable. Watch time also responds within the period, and that same-period part of a sustained loss lies outside the next-period estimand (Appendix B.1).

Protocol differences at higher shielding ratios. At the higher ratios of Table 1 two protocol differences contribute to the remaining gap. The two evaluations were five weeks apart, and the supply pool and its substitutes grew in between, so the later offline settlement faced stronger compensation. The 0–24 hour cold-start window carries about 20% of the online effect but is not fully covered by the offline settlement logs. A dedicated offline measurement of that window recovers a 19.4% increment, matching the share of the online effect carried by that window.

Cross-scenario validation on live-streaming monetization. To demonstrate cross-domain generality, we extended the evaluation to a live spending monetization proxy. In a targeted intervention suppressing 5% of SA streamers, the platform's highest streamer tier, the online revenue change was -1.83% and the settlement-based offline estimate -1.64%. The two have the same sign and are 0.19 pp apart. This single comparison extends the set-level check to a sparse, high-value metric outside engagement.

Set-level compensation diagnostics. The substitution coefficient decreases from 7.16 under the incumbent boosting strategy to 1.99 under SCSV (Appendix L). The shielding comparison above yields a larger loss per unit of removed supply, and the B-to-C discount coefficient falls under value-guided selection (Eq. (73)). These diagnostics assess selection by the settled credit and the prior distilled from it, at the level of selected sets. They do not isolate the replaceability discount from the sign gate and the unit budget.

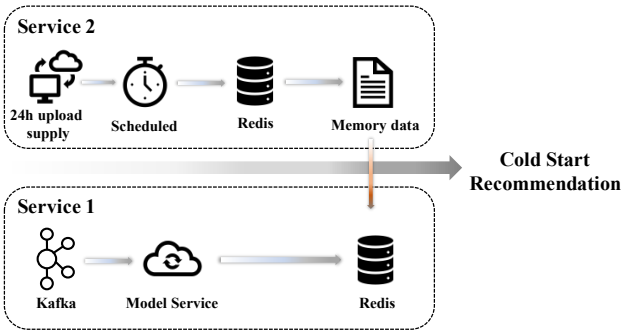


Figure 7: Nearline-to-online serving of SCSV for cold-start ramp-up. A nearline scorer writes per-item priors to Redis, and a scheduled job materializes them into the ranker's online feature (memory data) format for request-time use in Service.

K Deployment and Serving Details

We deploy SCSV as a nearline-to-online serving pipeline that exposes two per-item signals for newly uploaded inventory (Fig. 7), a prior value score used to shape cold-start exposure and an uncertainty signal used to enforce conservative ramp-up under cold-start and distribution shift. The deployment objective is twofold: (i) keep online ranking latency unchanged by performing all model inference off-path, and (ii) provide a monotone, auditable control surface (thresholds and budgets) so that exploration can be increased or rolled back safely. At request time, the online ranker performs only lightweight lookups and uses SCSV as a gating-and-ranking interface. It constructs a ramp-up candidate pool via a monotone eligibility gate that jointly enforces sufficient prior value and a bounded uncertainty proxy score:

$$\text{eligible}(i,t) \;\Leftrightarrow\; \pi_{i,t} \geq \tau_s \;\wedge\; u_{i,t} \leq \upsilon_s, \tag{74}$$

where (τ_s, υ_s) are surface-specific thresholds calibrated (via online experiments) to control ramp-up volume and tail-risk under shift. This gate is intentionally monotone. For a fixed candidate pool and fixed scores, increasing τ_s or decreasing υ_s can only shrink or leave unchanged the set eligible for SCSV-guided ramp-up. This monotonicity concerns eligibility. Its effect on exposure depends on downstream ranking and allocation. It provides a simple safety mechanism during rollout. Eligible items are injected into the candidate set (or a dedicated ramp-up bucket), and the prior value score is supplied to the existing ramp-up mechanism as an additional ranking feature, while downstream ranker constraints (e.g., diversity, safety, and policy constraints) remain unchanged.

The served signals align with the value-guided intervention construction in Section 2.3 and the evidential prior scorer in Appendix D. The supervised target distilled into the upload-time scorer is the substitution-corrected SCSV settlement label $y_{i,t} = T(R_{i,t}^{\mathrm{SCSV}})$ defined in Eq. (6) and Eq. (24), so the nearline output $(\pi_{i,t}, u_{i,t})$ in Eq. (31) is a serving-time approximation of the ex-post settled credit before next-period outcomes are observed. At request time, admitted items are boosted within the fixed ramp-up budget by the existing multi-factor ranker, which receives the prior value score as an additional feature. Eligibility permits consideration for ramp-up. It does not guarantee additional exposure, because the final ranking depends on the ranker's other inputs and on the allocation constraints. The risk-adjusted reference score $P_{i,t} = [\pi_{i,t} - \kappa\, u_{i,t}]_+$ of Eq. (32) is the risk-adjusted reference score of the framework in Section 2.3 and is not computed at serving time. The framework and the deployment share the value signal, not the decision rule. Proposition E.1 applies to the density-ordered selector of Section 2.3 (Appendix E). The deployed eligibility thresholds are empirically calibrated. They do not implement the positive score-based net-gain condition of Eq. (39). Operational safeguards include TTL-based eviction of stale priors, fail-closed defaults when KV entries are missing or invalid (no ramp-up granted), and surface-level traffic budgets that cap maximum ramp-up exposure, jointly playing the role of the finite attention budget C_t^+.

Deployment configuration For reproducibility, we summarize the deployment-time settings used in our online experiments: (i) The eligibility thresholds τ_s and υ_s are calibrated per surface to approximately the 80th percentiles of the historical π and u distributions on that surface. (ii) The traffic budget C_t^+ is set to 10% of total ramp-up impressions per surface. (iii) For ramp-up

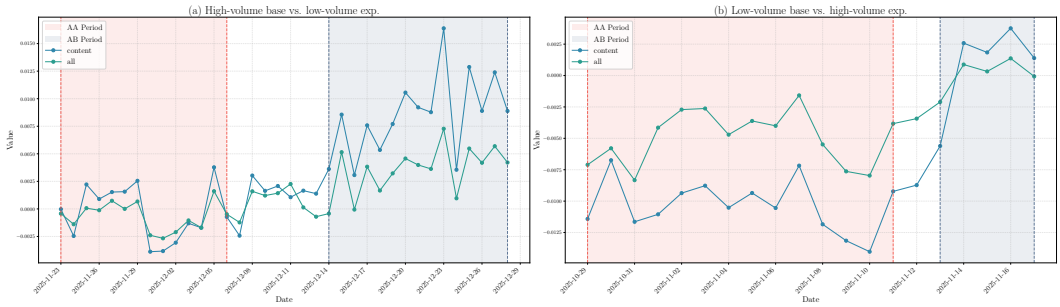


Figure 8: Difference-in-differences trend of the online effect. The treatment effect is estimated by comparing the experiment-control gap before and after SCSV-guided ramp-up.

bucket sizing, each experiment / control bucket receives 5% of total surface traffic, and the ramp-up window spans 7 consecutive days before evaluation. Exact numerical values are anonymized due to commercial confidentiality and will be released in the public artifact upon acceptance.

L Online Efficiency Interpretation and Case in Production Deployment

This appendix provides additional diagnostics for the production deployment in Section 4. The main text reports the core selection-efficiency and positive-work ROI results. Here we provide the difference-in-differences trend, pathway breakdown, cold-start liquidity analysis, traffic-scale robustness, and substitution-efficiency diagnostics.

Difference-in-differences trend. The online intervention effect is estimated using a difference-in-differences design that compares the experiment bucket and the control bucket across a 14-day pre-treatment AA period and a 15-day post-treatment AB period. Users are randomized into buckets by user ID. The experiment bucket receives 5% of traffic, corresponding to at least 20 million daily active users. The treatment effect is computed as the change in the experiment-control gap from AA to AB:

$$\Delta_{\mathrm{DiD}} = (Y_{\mathrm{exp}}^{\mathrm{AB}} - Y_{\mathrm{base}}^{\mathrm{AB}}) - (Y_{\mathrm{exp}}^{\mathrm{AA}} - Y_{\mathrm{base}}^{\mathrm{AA}}).$$

Fig. 8 reports the corresponding online trend.

Inference. Standard errors are clustered at the user level, the unit of randomization, to account for correlation across repeated observations within a user. Confidence intervals are 95% and all tests are two-sided, on the relative-lift scale. Table 13 reports the full estimates.

Fine-grained pathway decomposition. The pathway analysis decomposes open-reason changes by interaction source and content cohort. Fig. 9 reports, for the core-supply cohort of the demand-side experiment, the relative change and p-value of each open-reason category with the strategy in effect, next to two reference columns. Share-driven open reasons rise by $+1.54\%$ ($p = 0.044$), the only entry of the figure that differs from zero at the 5% level. Revisit-driven open reasons rise by $+0.38\%$ ($p = 0.211$), the search-, push-, and follow-driven categories move within noise, and the two reference columns show no change distinguishable from zero. Separately, the main production experiment reports a $+1.22\%$ increase in share-driven open reasons, a $+4.17\%$ increase in revisit-driven open reasons, and 78.51% of the observed variance in total open reasons explained by share plus revisit. These three figures are not read from Fig. 9.

Table 13: Full inference for the production DiD (relative-lift scale).

Value Type	Weekly Gain	Clustered SE	95% CI	p
View Time	$+0.053\%$	0.018%	$[+0.018\%, +0.088\%]$	0.003
Open Reasons	$+0.363\%$	0.155%	$[+0.059\%, +0.667\%]$	0.019
Retention	$+0.037\%$	0.016%	$[+0.006\%, +0.068\%]$	0.021
Revenue	$+0.025\%$	0.014%	$[-0.002\%, +0.052\%]$	0.074

Core Supply						
Open Reason	Strategy in effect		Strategy didn't work		Hsitory	
	Relative gain	P value	Relative gain	P value	Relative gain	P value
Search-driven	-1.72%	0.336	1.63%	0.310	-0.23%	0.162
Push-driven	0.05%	0.919	-0.06%	0.857	-0.16%	0.611
Share-driven	**1.54%**	**0.044**	-0.05%	0.926	0.90%	0.160
Follow-driven	-0.88%	0.305	0.15%	0.861	0.06%	0.934
Revisit-driven	0.38%	0.211	-0.12%	0.629	-0.33%	0.112

Figure 9: Pathway breakdown by open-reason category and content cohort in the demand-side experiment. The figure reports the relative change and p-value of each category for the core-supply cohort with the strategy in effect, next to two reference columns.

Both breakdowns place the gain in active-intent behaviors, explicit endorsement and active return, rather than in passively consumed content. In the demand-side experiment only the share-driven effect is individually significant.

Cold-start liquidity. SCSV also improves early-stage supply liquidity. The volume of high-UE breakout videos, defined as items surpassing 50k views with high satisfaction, increases by $+1.099\%$, and their proportion in the ecosystem increases by $+0.598\%$. This indicates that SCSV directs ramp-up traffic toward undersupplied high-quality inventory rather than further saturating already visible head supply.

Traffic-scale robustness. A secondary traffic configuration compares a small experiment bucket against a large base bucket. The effect remains directionally consistent, including a $+2.88\%$ increase in share-driven open reasons. This provides robustness evidence that SCSV-guided positive supply work is not limited to a single bucket-size configuration.

Mechanism reading. The production gains are consistent with the central SCSV mechanism, under which value-guided intervention should allocate finite ramp-up traffic toward supply whose marginal contribution is difficult to compensate by substitutes. In the open-reason objective, positive-work ROI increases from 13.99% to 50.00%, corresponding to approximately $3.6\times$ higher conversion from supply-side intervention to system-level value. The substitution coefficient below reports the same measurement, the reciprocal of this ROI, and adds a random-boosting reference that falls between the two.

Substitution coefficient. Internal cannibalization is quantified using a substitution coefficient, defined as the ratio between supply-side attributed lift and realized net demand-side gain:

$$\chi_{\text{sub}} = \frac{\Delta B_{\text{attr}}}{\Delta U_{\text{net}}}.$$

Smaller values indicate higher marginal efficiency under substitution. On a given metric the coefficient is the reciprocal of the positive-work ROI of Table 4, which divides the realized net gain by the attributed supply-side lift, so the two report one measurement in two conventions. The compensation-aware SCSV strategy achieves a coefficient of approximately 1.99, compared with 7.16 for the existing fine-grained boosting baseline and 5.27 for the random baseline. The first two correspond, up to rounding, to the open-reason ROI of 50.00% and 13.99% in Section 4, and the random baseline to an ROI of about 19%. The $3.6\times$ ratio between the baseline and SCSV coefficients is the ROI gain reported there, not an additional result.

Offline-estimate fidelity under deployment. The aggregate offline SCSV estimate is 0.8693% against an online value of 0.8736%. Residual deviations in finer-grained comparisons are mainly attributable to the 0–24 hour cold-start window, which accounts for a substantial share of early volume but is not fully captured in offline settlement logs.

Observations. As illustrated in Fig. 10, high-score items (low substitutability) are often narrative-, knowledge-, or skill-driven with clear temporal progression and higher information density (e.g., story telling and commentary, gameplay analysis, and episodic daily-life clips). In contrast, low-score

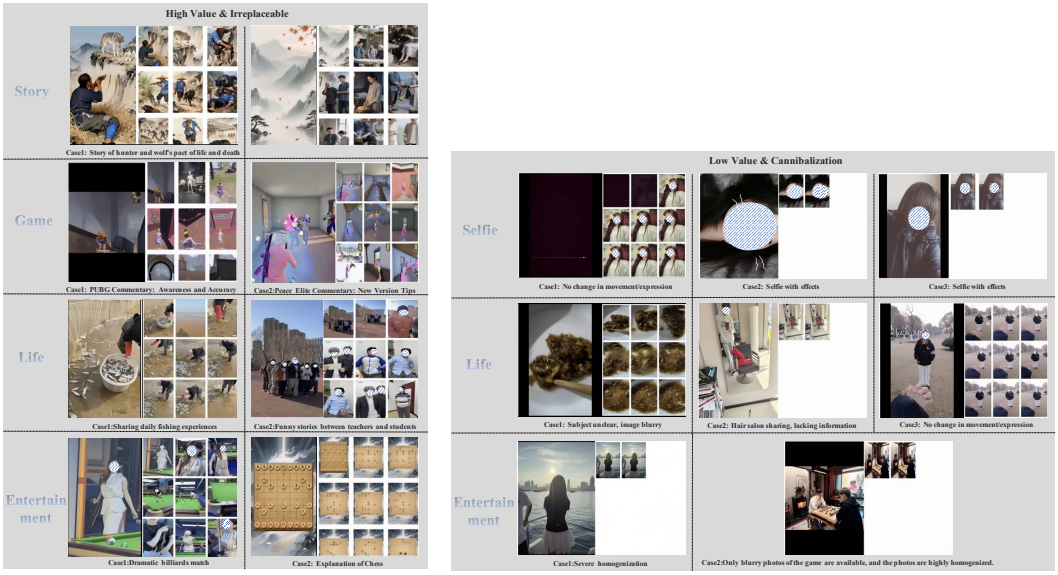


Figure 10: **Qualitative examples of the prior value score.** Left: high-score items (low substitutability) are typically narrative-, knowledge-, or skill-driven with clear temporal progression and higher information density (e.g., story telling, gameplay commentary, and episodic daily-life clips). Right: low-score items (high substitutability) are often homogeneous and low-information snippets (e.g., static mirror selfies, heavily filtered selfies, or blurry/poorly-framed daily-life shots). Thumbnails are anonymized.

items (high substitutability) are typically homogeneous and low-information snippets (e.g., static mirror selfies, heavily filtered selfies, or blurry/poorly-framed daily-life shots), which can be readily compensated by abundant substitutes.

M Extended Related Work

This appendix expands the related-work paragraph of the introduction.

M.1 Ecosystem objectives and supply-side incentives

Recommender systems are increasingly studied as multi-stakeholder ecosystems in which ranking policies shape both user utility and supply-side incentives. Prior work on creator dynamics, mechanism design, and welfare-aware recommendation shows that short-term engagement optimization can lead to inefficient long-term ecosystem outcomes (Ben-Porat et al., 2020; Boutilier et al., 2024; Mladenov et al., 2020). However, these studies do not provide an item-level value signal computable from production logs for estimating substitution-corrected marginal contribution or guiding positive supply intervention.

M.2 Substitutability and Competition Under Finite Attention

Theory has quantified how top-K ranking in competitive creator settings can induce substantial inefficiency and welfare loss through crowding-out effects (Yao et al., 2023). Empirical evidence from field experimentation further shows that recommendations can create true incremental consumption, but only after separating informational or exposure-driven gains from pure redistribution (Aridor et al., 2022). Zielnicki et al. (2025) make this separation structurally, fitting a discrete choice model to Netflix viewership and using recommender-induced variation to recover diversion ratios and the incremental engagement of the recommender as a whole. SCSV instead values individual supply items. At the modeling level, recent work has explored explicit representations of user choice and competition among items (Wang et al., 2023; Hazrati & Ricci, 2024), aiming to capture how users select among substitutes. He et al. (2026) instead constrain e-commerce uplift estimates to platform-

level consistency so that cross-seller cannibalization is not counted as incremental. Unlike user-choice models that require explicit substitution structure among items, SCSV captures substitution implicitly through an item-level replaceability discount calibrated against randomized shielding. It pairs this discount with a settlement budget that caps the credit of each attribution unit by its realized utility change.

M.3 Causal Identification and Evaluation Under Interference

Counterfactual evaluation from logged feedback underpins modern recommender systems, with learning-to-rank and off-policy estimators providing unbiased objectives under exposure bias when propensities are known or estimable (Schnabel et al., 2016; Joachims et al., 2017; Swaminathan & Joachims, 2015). However, these estimators assume SUTVA, which fails on recommender platforms where finite attention induces cross-item interference (Imbens & Rubin, 2015). The causal inference literature formalizes such spillovers with interference-aware estimands and identification, including cluster-based designs (Hudgens & Halloran, 2008; Aronow & Samii, 2017), and platforms operationalize these structures through experimental units such as ego-clusters (Saint-Jacques et al., 2019). For high-dimensional confounding, double/debiased machine learning provides orthogonal estimating equations that yield robust inference with imperfect nuisance fits (Chernozhukov et al., 2018). SCSV studies a structured setting in which interference is modeled through exposure reallocation under a specified serving policy. Within the response model, policy-specific compensation ratios summarize substitute recovery relative to nonzero gross contributions at a given reference allocation, utility metric, and removal dose. Their estimation relies on the modeling and calibration conditions of Section 2.2 and Appendix B.1. Conservation alone constrains feasible reallocations but does not identify these ratios. SCSV builds on these foundations by pairing lagged temporal differencing with DML-style deconfounding to estimate item-level attribution effects from logs, then settling these effects into substitution-corrected credit rather than assuming item independence. Finally, because interference makes purely offline validation fragile, evaluation commonly uses logged-bandit benchmarks and randomized exposure datasets (Li et al., 2010; 2011; Lefortier et al., 2016; Gao et al., 2022b). None pair production logs with randomized supply-shielding outcomes, which motivates our release of such a benchmark.

M.4 Value-guided Positive Supply Intervention under Finite Attention

Research on recommendation and ranking treats exposure as the allocation of a scarce resource. Fairness-of-exposure methods constrain how visibility is distributed across ranked items or groups (Singh & Joachims, 2018), while equity-of-attention approaches amortize attention across repeated rankings to align received attention with deserved relevance (Biega et al., 2018). In two-sided recommendation and marketplace settings, prior work further studies supplier exposure fairness across relevance and user satisfaction (Mehrotra et al., 2018; Patro et al., 2020; Mansoury et al., 2022). These studies establish that exposure is a constrained allocatable resource, but they neither optimize for incremental gains in marginal ecosystem utility nor account for the cannibalization cost induced when supporting one item displaces substitute supply. In contrast, our value-guided positive supply work consumes the substitution-corrected credit settled by SCSV as its decision signal and targets net ecosystem gain under finite attention.