More Ads, Happier Shoppers: Unified-Valuation Ad Allocation at Scale

Sergey Kolbin
skolbin@wayfair.com
Wayfair LLC
Seattle, WA, USA

Arushi Jain
ajain31@wayfair.com
Wayfair LLC
Boston, MA, USA

Manavender Malgireddy
mmalgireddy@wayfair.com
Wayfair LLC
Seattle, WA, USA

Kurt Zimmer
kzimmer@wayfair.com
Wayfair LLC
Boston, MA, USA

Masoum Mosmer
mmosmer@wayfair.com
Wayfair LLC
Mountain View, CA, USA

Patrick Phelps
pphelps@wayfair.com
Wayfair LLC
Boston, MA, USA

Abstract

Balancing short-term ad revenue with long-term user satisfaction is a central challenge in sponsored product search. We propose common currency, a unified ranking framework that merges sponsored and organic results within the same valuation. By translating predicted clicks and conversions into variable contribution dollars (VCD), both ad and organic items compete on a single, profit-based metric. We combine this objective with guardrails on relevance and ad load to preserve a high- quality customer experience. Our approach is implemented via a one-pass, order-preserving greedy merge, which we show is near-optimal in offline simulations despite the problem's inherent non-monotonicities. We also compare real-time and batch variants, finding that batch ranking recovers most of the profit uplift while incurring significantly less latency. In a six-week online experiment covering millions of keyword searches, our method allowed to expand ads footprint and increased ad revenue by 35% and, surprisingly, also improved conversion rates and long-term profit across ads and organic products together. These results suggest that when ads are relevant and well- spaced, elevating ad load can benefit both advertisers and end users, challenging the classic trade-off narrative. We discuss lessons learned from production deployment and outline directions for further enhancements.

Keywords

Sponsored Products Advertising, Ranking Optimization, Multi-objective Ranking, Profit Aware Optimization, Advertising Footprint, Profit-Relevance Tradeoff

ACM Reference Format:
Sergey Kolbin, Arushi Jain, Manavender Malgireddy, Kurt Zimmer, Masoum Mosmer, and Patrick Phelps. 2025. More Ads, Happier Shoppers: Unified-Valuation Ad Allocation at Scale. In *Proceedings of KDD '25*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

KDD '25, Toronto, Canada

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Sponsored product ads now shape the way millions of shoppers discover products online. Yet every additional ad competes with organic results for scarce real estate, forcing platforms to juggle three objectives that rarely align: (i) short-term ad revenue, (ii) short- and long-term retailer profit, and (iii) customer experience (e.g., measured by engagement and conversion rate CVR).

Status quo and blind spots. A common approach on e-commerce search pages is a static footprint. This design (a) leaves material revenue untapped whenever relevant ads run out of slots, (b) ignores the profit margin of what shoppers actually buy, (c) cannot adapt to user ad tolerance. Recent feed-style work, notably Yan et al. [13], tackles a different setting: an endless-scroll newsfeed. They blend two pre-ranked lists (ads and organics) with a single shadow-price parameter and enforce simple guardrails—top-slot and min-gap between ads. However, their model optimizes expected ad revenue and engagement, not profit metrics like variable contribution dollars (VCD) that matter in retail. It also omits grid-specific UX rules such as max consecutive ads or organic-spacing quotas.

Our approach. We formulate ad vs. organic allocation on product grids as a common currency optimization that prices every impression in dollar including ad revenue, variable contribution dollar (VCD - profit per sale), and a conversion guardrail:

$$u = \text{AdRevenue} + \text{pCVR}\big(\lambda + \text{VCD}\big),$$

subject to page-level relevance and fine-grained guardrails on density and spacing. Because those guardrails introduce only short-range dependencies, a single-pass *greedy merge* achieves $>99\%$ of the objective value of a more complex block-filling search.

Using two million historical searches we tuned the Lagrange multiplier λ and guardrails, then validated the policy in a six-week A/B test on U.S. traffic. Contrary to the classic "ad-fatigue" narrative, increasing the ads footprint simultaneously lifted ad revenue by 35%, and combined ads and organic VCD by 3% and CVR by 1%.

Our key contributions are:

(1) **Unified-valuation ranking for e-commerce grids.** We extend the ads-vs-organic framework of Yan et al. [13] to product grids, incorporate profit (VCD) and viewability discounts, and add block-length and spacing guard-rails tailored to product tiles.

(2) **Linear-time greedy merge with near-optimality in practice.** A one-pass algorithm attains ≥99 % of the objective and is lightweight enough for production deployment.
(3) **Large-scale evidence that more ads can be a *win–win.*** Offline simulations and a live A/B test show concurrent lifts in Ad Revenue, VCD, and CVR—demonstrating that higher ad density, when relevance-constrained, can enhance both monetization and shopper experience.
(4) **Bridging science and production constraints.** We quantify the trade-off between real-time and weekly batch computation, showing that cached, keyword-level footprints recover most of the ad revenue upside with no additional serving latency.

2 Related Work

Balancing sponsored and organic content is a classical multiobjective optimization problem. Below we outline two research strands and position our "common-currency" footprinting work within that context.

Constrained and multi-objective optimization for ad insertion . Google's long-term experiments showed that short-run revenue lifts can erode lifetime value, motivating guard-rails on ad load and relevance [7]. A decade later, LinkedIn operationalized this idea at feed scale: a post-ranking merge maximizes revenue subject to an engagement constraint and simple UX rules [13]. JD.com generalized the approach by assigning virtual bids to organic items so that ads and organics compete in a single auction [4]. We follow the same line, but adapt it to an e-commerce grid: (i) we price conversions in variable contribution dollars (VCD), (ii) introduce max consecutive ads and organic spacing guardrails suited to product tiles, and (iii) bridge science complexity with production constraints (show that a greedy or batch approximations can deliver most of the value).

Re-examining the "ad-fatigue" narrative. Google's long-horizon experiments famously showed that heavy ad loads can back-fire: extra sponsored links boosted short-run clicks yet reduced long-run usage once users learned to ignore clutter [7]. More recent evidence paints a subtler picture. Multi-objective ranking on Taobao raised both GMV and ad revenue in a 28-day test [14], and personalized ad-load policies on a 200-million-user social feed delivered simultaneous lifts in engagement and ad returns [11]. Optimized ad-pod scheduling on connected-TV streams likewise served additional ads while preserving viewer completion rates [8]. These studies suggest that when ads are relevant, well-spaced, and tailored to user tolerance, higher ad density need not erode organic engagement—in fact it can enhance overall value. Our own online experiment (§6) corroborates this emerging consensus: doubling ads footprint on grid search raised page-level CVR, VCD and long-term gross revenue and VCD, overturning the classic ad-fatigue concern in the context of Wayfair product search.

3 Problem Formulation

3.1 Candidate Sets

Let $O = \{1, \ldots, |O|\}$ be the organic candidates and $A = \{1, \ldots, |A|\}$ the sponsored candidates. For each $j \in O$ we have the predicted conversion (per impression) rate pCVR^o_j and per-order profit VCD^o_j. For each $k \in A$ we know

Table 1: Notation used in Section 3

Symbol	Description
$O,\ A$	sets of organic and ad candidates
$\mathrm{pCVR}^o_j,\ \mathrm{pCVR}^a_k$	predicted conversion rate per impression
$\mathrm{VCD}^o_j,\ \mathrm{VCD}^a_k$	profit per order (short-term)
ar^a_k	expected ad revenue (pCTR × CPC)
λ	shadow price of a conversion
δ_v	per-slot viewability discount
δ_s	per-slot offline slot-bias correction factor
C	minimum cumulative CVR
m	max consecutive ads
x	organic spacing multiplier after an ad block

$$\mathrm{ar}^a_k = \mathrm{pCTR}^a_k \cdot \mathrm{CPC}_k, \qquad \mathrm{pCVR}^a_k, \qquad \mathrm{VCD}^a_k,$$

where ar^a_k is the expected ad revenue in CPC model and is the product of predicted click-through rate pCTR^a_k and the clearing price CPC_k. Notice that our auction incorporates the floor price which bounds CPC_k from below.

Although the auction ranks ads by maximum bid, our page-level optimization uses the clearing price CPC_k instead. This aligns with Wayfair profit, removes the incentive to inflate bids, and follows the utility-vs-value perspective of [2, 6].

3.2 Page Variables

We need to fill N product slots in search results. A binary vector $\mathrm{y} = (y_1, \ldots, y_N) \in \{0, 1\}^N$ marks ads ($y_i = 1$) and organics ($y_i = 0$). Define

$$s(i, a) = \sum_{k=1}^{i} y_k, \quad s(i, o) = \sum_{k=1}^{i} (1 - y_k),$$

so $s(i, a)$ (resp. $s(i, o)$) is the running index of the ad (resp. organic) item placed in slot i.

3.3 Common-Currency Utility

We score each impression so that all terms can be expressed in currency (e.g. USD):

$$u^a_k = \mathrm{ar}^a_k + \mathrm{pCVR}^a_k\big(\lambda + \mathrm{VCD}^a_k\big), \qquad u^o_j = \mathrm{pCVR}^o_j\big(\lambda + \mathrm{VCD}^o_j\big).$$

The common currency formulas for ad and organic products are the same while recognizing that organic products do not generate ad revenue. The shadow value λ converts orders into the same units as profit and ad revenue (e.g. USD). With a slightly abuse of notation, we include λ directly into utility, see below for further details.

3.4 Objective

For each user search the objective is as follows:

$$
\begin{aligned}
\max_{\mathbf{y}\in\{0,1\}^N} \quad & \sum_{i=1}^{N} \delta_v^{\,i} \cdot \left(y_i \cdot u^a_{s(i,a)} + (1-y_i) \cdot u^o_{s(i,o)}\right) \\
\text{s.t.} \quad & \sum_{i=1}^{N} \left[y_i \cdot \mathrm{pCVR}^a_{s(i,a)} + (1-y_i) \cdot \mathrm{pCVR}^o_{s(i,o)}\right] \ \geq \ C \quad \text{(Rel.)} \\
& \text{max ad block length} \ \leq m \quad \text{(LG-1)} \\
& \text{after ad block of length } b \text{ insert } \lfloor x \cdot b \rfloor \text{ organics} \quad \text{(LG-2)}
\end{aligned} \tag{1}
$$

$\delta_v \in (0,1]$ models viewability decay with scroll depth. Constraint (Rel.) enforces a minimum page CVR and allows us to determine the shadow value of an order λ (shared across all user searches) through offline simulations in Section 5. LG-1 caps any contiguous run of ads at m items ($m \leq 4$). LG-2 requires at least $\lfloor x \cdot b \rfloor$ organic results after an ad block of length b, rounded down for fractionals, with $x \in [0.5, 2]$.

Deduplication. If a product appears in both candidate sets, we keep its first occurrence and drop the duplicate; the same rule is applied in offline simulations.

Order-preserving fill vs. footprint stability. Our solver keeps the program-internal order produced by the Ads and Organic rankers as in [13]. This modularity means each team can evolve its model independently, but it also ties total footprint to upstream rank changes (e.g., if one program wants to boost certain products, its average pCVR declines reducing footprint). We tested an order-agnostic version that first re-ranks all product candidates based on common currency and only then decides the footprint; it hurt key metrics in offline simulations due to the wedge in footprint determination and actual SKUs placed. We therefore accept the order-preserving trade-off and plan to manage long-term stability by tuning λ, tightening guardrails, or rolling out boosting changes symmetrically across programs.

4 Merge Algorithms

Below we present two constructive heuristics that build a merged page slot by slot from the pre-ranked Ad and Organic lists. Even though we decided to proceed with Greedy in practice, we evaluated that the loss is small relative to a more optimal approximate solution.

4.1 Greedy merge (Algorithm 1)

Algorithm 1 is a single-pass $O(N)$ merge identical in structure to the rule used by [13]. Function `GuardrailsOK`(L, m, x) returns true if appending one more ad to the currently merged list L keeps the current contiguous ad block $\leq m$ and we already placed at least $\lfloor x \cdot b \rfloor$ organics after the last ad block.

4.2 Block-filling search (Algorithm 2)

Why greedy is sub-optimal. Each program's list is internally monotone in its own score, yet the common-currency utilities $\{u^a, u^o\}$ need not be monotone. Consider filling $N = 3$ slots with common currency utilities for ads as $(0.02, 0.06, 0.03)$ and organic as $(0.04, 0.03, 0.02)$. Greedy places the 0.04 Organic first for the total of 0.04+0.03+0.02 = 0.09. The optimal layout starts with two Ads and achieves 0.02+0.06+0.04 = 0.12. Greedy's myopic, slot-by-slot choice cannot foresee that a slightly worse item now may unlock a much better one next.

Algorithm 1: GREEDY MERGE

Input: ranked organics $L^o[1 \ldots |O|]$, ranked ads $L^a[1 \ldots |A|]$; max ad-block length m, spacing factor x; target page length N
Output: merged list $L[1 \ldots N]$

$j_o \leftarrow 1,\ j_a \leftarrow 1$
while $j_o + j_a \leq N$ **do**
  if $j_a \leq |A|$ ***and*** $u^a_{j_a} > u^o_{j_o}$ ***and*** `GuardrailsOK`(L, m, x) **then**
    L.append$(L^a_{j_a})$;
    j_a++
  else
    L.append$(L^o_{j_o})$;
    j_o++
  end
end
return L

BlockFill approach. We propose a more optimal approach that would result in an optimal solution for the example above. The approach considers all possible combinations of ad block lengths for each decision whether to put an ad and thus also has a linear complexity of $O(Nm)$ (with $m \leq 4$ in practice).

Algorithm 2: BLOCKFILL (m-slot look-ahead)

Input: same inputs as Algorithm 1
Output: merged list $L[1 \ldots N]$

$j_o \leftarrow 1,\ j_a \leftarrow 1$
while $j_o + j_a \leq N$ **do**
  for $b \leftarrow 0$ **to** m **do**
    $U_b \leftarrow \sum_{t=0}^{b-1} \delta_v^{\,t}\, u^a_{j_a+t} + \sum_{t=0}^{m-b-1} \delta_v^{\,b+t}\, u^o_{j_o+t}$
    discard U_b if the block $(b,\ m-b)$ violates guardrails
  end
  $b^\star \leftarrow \arg\max_b U_b$
  if $b^\star > 0$ *(can put ads preserving guardrails)* **then**
    L.append$\left(L^a[j_a : j_a + b^\star - 1]\right)$
    $j_a \leftarrow j_a + b^\star$
  else
    L.append$(L^o_{j_o})$
    j_o++
  end
end
return L

Empirical comparison. We compared two approaches in offline simulations (not reported for brevity) and observed that Greedy merge consistently achieves ≥ 99% of BlockFill objective across guardrail and constraint combinations. The negligible gain does not justify the added complexity, so we proceeded with Greedy merge for further simulations and tests. We leave finding the exact solution for future work.

5 Offline Simulation Study

Offline simulations let us stress-test candidate footprint policies, select the Lagrange multiplier λ, and pick customer-experience guardrails without touching live traffic.

5.1 Data & Pre-processing

Raw data. We use a sample of 4M user keyword searches on Wayfair U.S. traffic for the two weeks before the A/B launch. We split these searches into the training set (first week) to generate the footprints and evaluation set (second week). Each user search (e.g., "red sofa") yields several hundred ad *and* organic candidates per query, each with:

- predicted conversion rate pCVR, same deep learning model applied to both ads and organic products
- variable contribution dollars VCD,
- for ads only: predicted click-through rate pCTR and cost-per-click CPC outcome from the advertising auction.

Keyword coverage. We retain and report the results for the evaluation set based only on the keywords that are present in the training set - this helps us approximate how the cached (batch) footprint would perform in production. Replaying a cache that is one week old introduces negligible staleness in practice. For the long tail of new user searches that are not covered by keyword based batches, we developed an approach relying on a higher level aggregation (e.g., product class instead of keyword) and that was not included in the original test, see Conclusion for further details.

Position-bias correction. Because the models are position-agnostic, we down-weight engagement predictions by a slot-specific factor, e.g., pCVR ← pCVR $\cdot\delta_s^i$ with $\delta_s \in (0, 1)$. Without this adjustment, bottom-of-page pCTR and pCVR would be overstated more than 2×, distorting the offline projections. Notice that δ_s^i is the correction for model bias and would not be needed had the models contained slot as a feature, while δ_v^i is an independent discount that captures user behavior.

5.2 Metrics

For both offline simulations and online tests we track two complementary metric groups that are expected to capture both advertising program growth and overall company goals including user experience. We do not simulate ROAS and gross revenue, but report those outcomes in the online test.

Advertising key metrics.

- **Ad revenue**: $\widehat{\mathrm{Rev}} = \sum_i \delta_v^i \cdot \mathrm{pCTR}_i \cdot \mathrm{CPC}_i$
- **Ad clicks**: $\widehat{\mathrm{Clicks}} = \sum_i \delta_v^i \cdot \mathrm{pCTR}_i$
- **Footprint**: number of ad slots on a page = $\sum_i y_i$ (distinct from impressions because many users do not scroll)

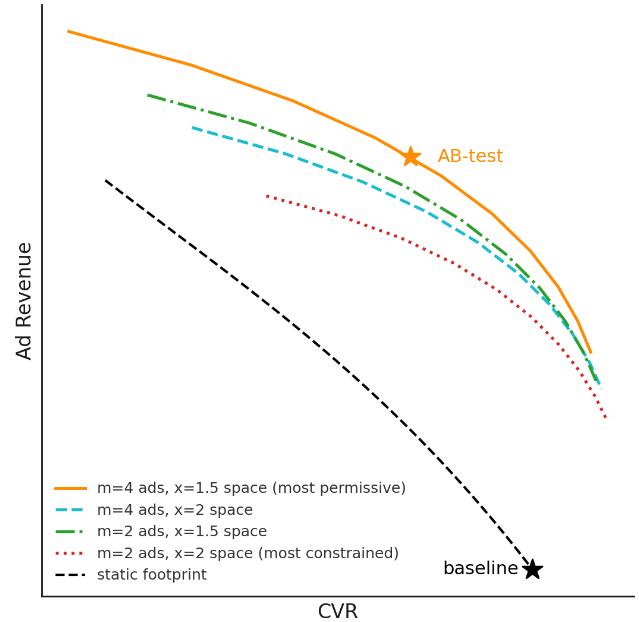

Figure 1: Guardrail trade-off. Each colored curve is the *Ad-Revenue–CVR* frontier for a guardrail pair (m, x); the dashed line is the static baseline. Stars mark the production policy (black) and the A/B allocation (orange). Units suppressed for commercial sensitivity.

Advertising + organic joint metrics.

- **CVR**: $\widehat{\mathrm{CVR}} = \frac{1}{N} \sum_i \delta_v^i \cdot \mathrm{pCVR}_i$
- **VCD**: $\widehat{\mathrm{VCD}} = \sum_i \delta_v^i \cdot \mathrm{pCVR}_i \cdot \mathrm{VCD}_i$

5.3 Selecting the policy to test

Guardrail search. For each combination of *max ad–block length* $m \in \{2, 4\}$ and *spacing multiplier* $x \in \{1.5, 2.0\}$ we sweep the Lagrange multiplier λ and record the resulting *Ad-Revenue*–vs–*CVR* frontier. Figure 1 shows four such frontiers: the looser guardrails (m=4, x=1.5) dominate the stricter (m=2, x=2) alternative across the entire curve, so we adopt the orange policy for the A/B test (orange star) with specific λ value chosen based on the business trade-off between CVR and ad revenue. Notice this value of λ is also high enough to be consistent with the long-term impact of a generated conversion on company's long-term VCD (purchases today stimulate purchases in the future).

Production baseline. The black star marks the live static-footprint policy (ads fixed in slots 2, 4, 8, 9, 14, 15). Extending that static list would shift the dashed line up (revenue) and left (lower CVR), but the incremental gains are far smaller than those available from dynamic footprints, so we did not pursue larger static layouts.

5.4 Offline simulation: results & trade-offs

In the previous section we used simulations to confirm that we are more likely to succeed with a dynamic footprint approach and selected a specific set of guardrails. We now perform additional simulations to achieve two additional goals: 1) decide whether we want to build a real time system vs. using a batch approach and 2) determine if we want to test a version without VCD.

Table 2: Offline-simulation lifts vs. a static-footprint baseline (positive = better). RT = real-time ranker, Batch = daily batch ranker, NoVCD = value function without the VCD term.

	RT–VCD	RT–NoVCD	Batch–VCD	Batch–NoVCD
Panel A: Advertising Metrics				
Ad revenue	38.7	48.1	38.1	46.9
Ad clicks	39.9	46.0	39.4	45.5
Avg. footprint lift	99.3	104.3	103.6	108.3
Panel B: Advertising + Organic Metrics				
CVR	0.3	0.5	- 0.8	- 0.8
VCD	3.0	- 0.6	0.7	- 1.0

Real-time (RT) vs. Batch - Motivation. Algorithms described in Section 4 determine optimal footprints given data inputs per each user search. If we want to apply those algorithms as is in production, we then would need a real time implementation as we will have to rely on scoring model outputs such as pCTR and pCVR. Unfortunately, real time solutions are usually more complex to implement in practice as that would require more synchronization between the ads and organic systems to combine both ads and organic score inputs inside the footprint system. Real-time solutions can also increase latency due to both synchronization and the need to fetch additional inputs such as bids or VCD. Therefore we consider an alternative, a batch solution in which we precompute footprints for each keyword, taking the footprint of the median length over all user searches for that keyword in the past week. Notice that we can compare the two approaches in offline simulations as the required inputs are logged in our system per each user search.

Real-time (RT) vs. Batch - Outcome. Table 2 reports the outcome of our simulation analysis and shows that RT can generate more ad revenue (+0.6-1.2%), higher CVR (+1.1-1.3%) and VCD (+1.6-2.3%) while requiring lower ad footprint. While these differences are important, we decided to implement the first online test using the Batch approach due to faster time to production and still noticeable ad revenue opportunity.

Effect of the VCD term. When we include the VCD component (RT-VCD and Batch-VCD), we observe higher projected VCD (+1.7-3.6%) but lower ad revenue (-8.8%-9.4%), so we test both the VCD and no-VCD settings in production. We include the no-VCD version because of higher ad revenue and we generally expect that ad revenue simulation predictions would be more accurate (e.g., ad revenue is based on clicks that are easier to predict due to more click data).

Ad revenue vs. footprint. Projected ad revenue increases are smaller than footprint increases - while we can allow ads to take over more slots, those slots are often at the bottom of the page and so are viewed and clicked less often by the users.

Reproducibility. All simulations run in Google BigQuery with the ranking logic expressed as embedded JavaScript UDFs, enabling end-to-end offline simulation reruns from a single SQL script.

6 Online Experiments

We ran a six-week A/B test on Wayfair's U.S. keyword search traffic, comparing two batch-computed policies (Batch–VCD and Batch–NoVCD) with the production static footprint (all variants received 1/3 of traffic randomized by user id hash). All reported lifts are percentage changes versus baseline.

Additional metric details. In the online A/B results we also report ROAS (Return on Ad Spend) and gross revenue - these guardrail metrics were not included in the offline simulations. While we report impression-level CVR, VCD, and gross revenue metrics in Panel B, we also confirmed that the results are consistent on page or user session levels. The online gains mirror the direction of our offline projections but are modestly smaller, as simulation estimates do not fully capture serving-time noise and model error. To tighten this offline–online gap, we are evaluating counterfactual-estimation approaches such as Nguyen et al. [9] for the future.

Table 3: Online A/B percentage lifts; bold values are significant at $p < 0.05$ (two-sided t-test).

	Batch–VCD	Batch–NoVCD
Panel A: Advertising Metrics		
Ad revenue	**28.3**	**35.0**
Ad clicks	**28.3**	**31.7**
Avg. footprint lift	**78.3**	**81.7**
ROAS	**7.4**	**5.9**
Panel B: Advertising + Organic Metrics		
CVR	- 0.7	**1.0**
VCD	1.2	**2.7**
Gross revenue	0.4	**2.6**
Panel C: Long-Term Site-Wide Projections		
VCD	0.3	**0.6**
Gross revenue	0.3	**0.6**

Win–win lift. Both treatments increased ad revenue, and the Batch–NoVCD arm additionally produced statistically significant lifts in page-level CVR, VCD, gross revenue, and ROAS. Prior work shows that heavy or poorly targeted ad loads can depress engagement and retention [3, 5]. Here we moved from six ads per page to a still-conservative footprint of average 12 out of 48 slots per page chosen based on expected common-currency utility under pCVR guardrails. For some keywords we showed fewer ads than before (e.g. when ads had much lower pCVR than organic products). The extra ads were often more relevant than the displaced organic results, consistent with advertisers exploiting keyword-level gaps to surface high-intent products that organic ranking misses. Our findings align with recent evidence that carefully allocated, high-relevance ads can benefit both users and platforms [11], demonstrating that ad load need not be a zero-sum trade-off when relevance constraints are enforced.

Batch–NoVCD outperforms the VCD-aware variant. The simpler Batch–NoVCD arm delivered significantly higher ad clicks, ad revenue, and page-level CVR/gross revenue, and even showed a (non-significant) lift in VCD—paradoxically surpassing the variant that optimizes VCD directly. This underscores a pragmatic lesson:

enriching an objective with extra terms does not guarantee better business outcomes. A likely culprit is estimation noise: predicting VCD is harder than predicting clicks or conversions, so errors propagate into the footprint and erode gains. Offline–online gaps of this sort are well documented in large-scale ad systems [10]. Closing them will require higher-fidelity VCD models or uncertainty-aware objectives, which we leave for future work.

Long-term durability. Ad-fatigue can reverse early gains [7], so we ran both variants through Wayfair's Long-Term Impact Estimation Platform, which extrapolates effects via the surrogate-index method of Athey et al. [1]. Using 42 days of experiment data, the platform projects metrics out to 168 days across the entire site (i.e., beyond keyword search covered by our experiment). The forecasts show positive lifts in both VCD and gross revenue; only the Batch–NoVCD arm is statistically significant, but the direction is consistent for both. Hence the short-run revenue uplift is unlikely to be clawed back by latent ad-fatigue and is expected to translate into durable profit.

7 Conclusion and Future Work

Conclusion. We deployed a unified–valuation "common currency" ranker that prices ads and organics in the same profit-based units, then merges the two lists via a one-pass greedy algorithm. We evaluated both real time and batch version and demonstrated the relative efficiency of a simpler batch approach in offline simulations, providing a path to apply our approach under a variety of ad and organic system designs (e.g., whether or not it's easy to connect ads and organic score inputs in real time). In production, the system allowed scaling footprint expansion and raised ad revenue by 35 % while also lifting page-level CVR and long-term profit, providing another example that carefully chosen ads can break a common revenue–UX trade-off.

Future work. Next steps fall into four tracks. (1) Broader reuse: our batch framework already supports footprints for unseen keywords by applying footprints fitted for predicted class of the user search and can further be applied in other context such as carousels in product pages based on anchor product. (2) Real time ranking: offline experiments suggest additional CVR and VCD gains from recomputing footprints on the fly; implementing this low-latency pipeline is our main engineering goal. (3) Stronger profit signals: we can try refining our VCD modeling so that profitability can become a direct optimization objective. (4) Developing more personalized ad supply focusing on user preferences for ads (e.g., such as in [12]).

References

[1] Susan Athey, Raj Chetty, Guido W. Imbens, and Hyunseung Kang. 2019. *The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely.* NBER Working Paper 26463. National Bureau of Economic Research. http://www.nber.org/papers/w26463 Revised August 2024.

[2] Santiago R. Balseiro, Kshipra Bhawalkar, Zhe Feng, Haihao Lu, Vahab Mirrokni, Balasubramanian Sivan, and Di Wang. 2024. A Field Guide for Pacing Budget and ROS Constraints. In *Proceedings of the 41st International Conference on Machine Learning (ICML '24) (Proceedings of Machine Learning Research, Vol. 235).* PMLR, Vienna, Austria, 2607–2638. https://proceedings.mlr.press/v235/balseiro24a.html arXiv:2302.08530.

[3] Andrei Z. Broder, Massimiliano Ciaramita, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, Donald Metzler, Vanessa Murdock, and Vassilis Plachouras. 2008. To Swing or Not to Swing: Learning When (Not) to Advertise. In *Proceedings of the 17th ACM International Conference on Information and Knowledge Management (CIKM '08).* 1003–1012.

[4] Carlos Carrion, Zenan Wang, Harikesh S. Nair, Xianghong Luo, Yulin Lei, Peiqin Gu, Xiliang Lin, Wenlong Chen, Junsheng Jin, Fanan Zhu, Changping Peng, Yongjun Bao, Zhangang Lin, Weipeng Yan, and Jingping Shao. 2023. Blending Advertising with Organic Content in E-Commerce: A Virtual Bids Optimization Approach. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI '23).* AAAI Press, Washington, DC, 15476–15484. doi:10.1609/aaai.v37i13.26835

[5] Ali Goli, David H. Reiley, and Hongkai Zhang. 2025. Personalizing Ad Load to Optimize Subscription and Ad Revenues: Product Strategies Constructed from Experiments on Pandora. *Marketing Science* 44, 2 (2025), 327–352. doi:10.1287/mksc.2022.0357

[6] Negin Golrezaei, Ilan Lobel, and Renato Paes Leme. 2021. Auction Design for ROI-Constrained Buyers. In *Proceedings of the Web Conference 2021 (WWW '21).* ACM, Ljubljana, Slovenia, 1007–1018. doi:10.1145/3442381.3449841

[7] Henning Hohnhold, Deirdre O'Brien, and Diane Tang. 2015. Focusing on the Long-Term: It's Good for Users and Business. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15).* ACM, Sydney, Australia, 1849–1858. doi:10.1145/2783258.2788583

[8] Niranjan Kumawat, Manu Vajpai, and Nitish Varshney. 2022. Programmatic Optimization of Ad Pods for Maximizing Consumer Engagement and Revenue. In *Proceedings of the 16th ACM AdKDD Workshop on Advertising and E–Commerce (AdKDD '22).* Washington, DC, USA. https://adkdd.org/2022/papers/adkdd22-kumawat-programmatic.pdf Extended abstract.

[9] Phuong Ha Nguyen, Djordje Gligorijevic, Arnab Borah, Gajanan Adalinge, and Abraham Bagherjeiran. 2023. Practical Budget Pacing Algorithms and a Simulation Test-Bed for eBay Marketplace Sponsored Search. In *Proceedings of the 2023 ACM AdKDD Workshop on Advertising and E–Commerce.* ACM, Long Beach, CA, USA. Extended abstract.

[10] Cathy Qian, Aayush Mudgal, Yinrui Li, Jinfeng Zhuang, Shantam Shorewala, Yiran Zhao, and Harshal Dahake. 2024. Handling Online–Offline Discrepancy in Pinterest Ads Ranking System. https://medium.com/pinterest-engineering/handling-online-offline-discrepancy-in-pinterest-ads-ranking-system-8fd662da4c2d. Pinterest Engineering Blog, accessed 18 May 2025.

[11] Hitesh Sagtani, Madan Gopal Jhawar, Rishabh Mehrotra, and Olivier Jeunen. 2024. Ad-Load Balancing via Off-Policy Learning in a Content Marketplace. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM '24).* ACM, Mérida, México, 586–595. doi:10.1145/3616855.3635846

[12] Wei Shi, Chen Fu, Qi Xu, Sanjian Chen, Jizhe Zhang, Qinqin Zhu, Zhigang Hua, and Shuang Yang. 2024. Ads Supply Personalization via Doubly Robust Learning. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)* (Boise, ID, USA) *(CIKM '24).* Association for Computing Machinery, New York, NY, USA, 4874–4881. doi:10.1145/3627673.3680035

[13] Jinyun Yan, Zhiyuan Xu, Birjodh Tiwana, and Shaunak Chatterjee. 2020. Ads Allocation in Feed via Constrained Optimization. In *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '20).* ACM, Virtual Event, 3386–3394. doi:10.1145/3394486.3403391

[14] Yukun Zheng, Jiang Bian, Guanghao Meng, Chao Zhang, Honggang Wang, Zhixuan Zhang, Sen Li, Tao Zhuang, Qingwen Liu, and Xiaoyi Zeng. 2022. Multi-Objective Personalized Product Retrieval in Taobao Search. In *arXiv preprint arXiv:2210.04170.* doi:10.48550/arXiv.2210.04170