MATEL: MULTI-ARM TREATMENT EFFECT LEARNER FOR E-COMMERCE PACKAGING OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Automated packaging selection at e-commerce scale assigns a pack type to every customer order, yet these decisions remain largely heuristic and cannot reason jointly about competing objectives: damage protection, material waste, labor, and transportation efficiency. We frame packaging selection as a multi-armed causal optimization problem and introduce the Multi-Arm Treatment Effect Learner (MATEL), a causal deep-learning algorithm that estimates the effect of each pack type on per-shipment damage rates from large-scale fulfillment data. MATEL advances causal effect estimation on four fronts: (1) a multi-armed causal prediction engine trained with self-calibrating loss and hybrid representation layers; (2) interchangeable state-of-the-art tabular backbones; (3) robustness to treatment sparsity, via renormalized pairwise-propensity weighting that keeps counterfactual contrasts bounded and the pairwise calibration doubly robust under overlap; and (4) transitive pairwise contrasts with bounded extrapolations for rarely-used pack types. Empirically, MATEL recovers the directional agreement of the observed pairwise treatment effect on 91% of focal pairs versus 62% for a double machine learning (DML) benchmark, while improving per-unit CATE directional agreement by $\geq$ 12 pp. Its estimation keeps every per-unit effect bounded on the sparse arms where benchmarks diverge. BLP slope, GATES, and the doubly robust curve show MATEL is better calibrated for effect ranking than the benchmark. On synthetic data, MATEL achieves the lowest ATE error and best ranking against four benchmarks, and the lowest PEHE under weak and severe overlap. MATEL is deployed in production; a production-calibrated simulation estimates a 3.74 pp network-cost improvement and a 25% reduction in corrugated-box usage, reducing packaging waste and associated carbon emissions.

1 INTRODUCTION

Operational systems routinely choose one of several actions for each unit under cost, capacity, and safety constraints. Such choices are fundamentally causal because an action's value is determined by the counterfactual outcome it would produce, not only by outcomes observed under a legacy policy. The right architecture feeds causal effect estimates into a constrained optimizer that selects the best feasible action per unit. Learning those effects becomes difficult when the decision context is high-dimensional and treatment observations are sparse. We address these challenges by combining causal inference with scalable deep learning and develop the framework through a large-scale packaging-selection problem. The design principles here apply well beyond the packaging setting.

Packaging selection determines which pack type is used for each shipment, from rigid corrugated boxes to flexible mailers (Figure B.1). This single choice affects several operational objectives: how well a product is protected in transit, and therefore the rates of damage *concessions* reported by customers and *problem-solve* (PS) events detected in fulfillment facilities; how much packaging weight and volume the shipment carries, affecting material use, transportation cost, and carbon emissions; and how quickly the shipment can be packed, affecting labor throughput. Existing decisions are largely rule-based and do not quantify these trade-offs for each product and package type. Consequently, they cannot determine whether the protection gained from a rigid package justifies its additional material, labor, and transportation costs.

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

The central difficulty is not predictive accuracy but *causal* structure. A supervised model can predict observed outcomes on pack types, but a model optimized for factual fit cannot recover the treatment effect $\tau_{a,b}(x) = \mathbb{E}[Y(a) - Y(b) \mid X{=}x]$ between pack types for the same product. Assignment is confounded since fragile, high-value items are routed to corrugated boxes by design, so a low-loss predictor attributes a box's protective value to the items selected into it, and its implied counterfactuals can even invert the true ordering. Features that correlate to the outcome and treatment can confound the relationship between treatment and outcome. These are the gaps between predictive and causal machine learning Shi et al. (2019); Shalit et al. (2017); Künzel et al. (2019); Chernozhukov et al. (2018). In our setting, the treatment is not binary but spans K pack types, so the estimand is the full set of pairwise contrasts $\tau_{a,b}$, which must be estimated jointly and remain mutually transitive. Compounding this, treatment assignment is sparse: for many products, certain pack-type cells are thinly observed or never observed at all. This is the regime where classical inverse-propensity and DML estimators fail to deliver bounded, unbiased estimates.

We introduce the **Multi-Arm Treatment Effect Learner (MATEL)**, a causal deep-learning framework designed for this setting. MATEL combines a flexible representation backbone with K arm-specific Tweedie outcome heads and a separate propensity encoder over identifying confounders. For each treatment pair, it renormalizes the two arm propensities into a conditional pairwise propensity and performs doubly robust targeted calibration using observations from the focal pair. It then reconciles the independently calibrated per-unit contrasts against the absolute counterfactual levels, producing a coherent and non-negative set of arm outcomes whose differences are transitive by construction. The resulting estimates feed a constrained optimizer that selects a feasible package type for each shipment while balancing predicted damage against operational costs.

Contributions. We have the following main contributions in this paper:

- **Multi-arm treatment-effect learner with a self-calibrating objective.** MATEL generalizes targeted neural estimation to K treatments through arm-specific Tweedie heads for heavy-tailed outcome, and a two-stage regularization procedure for multi-arm settings.
- **Debiased estimations under sparsity.** MATEL provides bounded estimations for sparse treatments where positivity-violation degrades traditional methods.
- **Flexible representation learning with a hybrid propensity design.** The architecture supports interchangeable tabular backbones, including ResNet-PLE and FT-Transformer, as well as multimodal feature augmentation. A separate propensity encoder uses only confounders, preventing outcome-only features from destabilizing treatment weights.
- **Reconciliation: transitive treatment effect estimations.** A reconciliation procedure maps independently calibrated pairwise effects into one set of per-unit counterfactual outcomes whose differences are transitive. The estimands are paired with influence-function-based uncertainty quantification, with variance floors for rarely-observed treatment cells.
- **End-to-end decision system that generates major real-world impact.** MATEL outperforms benchmark causal estimators in our comprehensive empirical studies. We couple MATEL with a multi-objective packaging optimization engine in production. The simulated production scenario reflects 3.74 pp entitlement across a large fulfillment network.

1.1 Related Work

Causal machine learning spans effect estimation, representation learning, and decision making (Kaddour et al., 2022), built on the propensity score framework (Rubin, 1974; Rosenbaum & Rubin, 1983). Modern heterogeneous-effect estimators include meta-learners, Bayesian outcome models, causal forests, and balanced representation learning (Künzel et al., 2019; Nie & Wager, 2021; Hill, 2011; Athey et al., 2019; Johansson et al., 2016; Shalit et al., 2017); DragonNet adds targeted regularization to align prediction with causal estimation (Shi et al., 2019), connecting neural representation learning with doubly robust AIPW, TMLE, and DML (Robins et al., 1994; van der Laan & Rubin, 2006; Chernozhukov et al., 2018). Most target binary treatment. Multi-arm extensions include Perfect Match's propensity-matched minibatch augmentation (Schwab et al., 2018), multi-valued DML (Xiang et al., 2025), treatment-balancing representations (Mondal et al., 2022), and HydraNet's multiclass propensity model with per-arm outcome heads (Velasco-Regulez & Cerquides, 2023). To our knowledge, no prior work jointly addresses multi-arm effect learning, sparse pairwise overlap, and per-unit reconciliation. MATEL targets this intersection. Appendix A details the benchmark selection in this study and its correspondence to the related estimator families.

2 PROBLEM SETUP

The core question MATEL answers is a counterfactual: *what DPMO would this product have incurred had it been shipped in pack type k?*

2.1 CAUSAL FORMULATION

We frame pack-type selection as a multi-arm causal inference problem under unconfoundedness. Each unit i is a product with covariates $Z_i = (W_i, X_i) \in \mathbb{R}^d$ over a hundred dimensions, spanning product attributes (Table B.1), packaging-relevant characteristics, operational features, co-shipment composition, geographical regions, and transportation routes (ground or air); post-treatment variables or mediators are identified and excluded from Z_i. We partition Z_i into confounders W_i (affecting assignment and outcome) and treatment-independent features X_i (affecting only the outcome), as in Figure B.2. W_i is augmented with multimodal text and image embeddings. The treatment $T_i \in \{1, \dots, K\}$ is the pack type assigned to i from a candidate eligibility set, with Box-A as the baseline arm b. For each unit and pack type k, the outcome $Y_i(k) \in \mathbb{R}_{\geq 0}$ is defects per million opportunities (DPMO) that i would have realized in pack type k, modeled separately for concession and PS. The outcome is zero-inflated and heavy-tailed, where most products see no damage.

Identification of $\mathbb{E}[Y(k)|W, X]$ requires three standard assumptions: (a) SUTVA, one shipment's damage does not depend on how others are packed; (b) conditional independence, $\{Y(1), \dots, Y(K)\} \perp\!\!\!\perp T \mid W, X$, stated on both W and X because that is the set the outcome heads read; and (c) positivity, $P(T = k|W = w) > \varepsilon$ for business feasible (w, k). The condition $X \perp\!\!\!\perp T \mid W$ is what lets the propensity encoder read W alone (Section 3.1), and we restrict X to pre-treatment variables. Positivity is the assumption empirically strained on sparse arms; Section 3.3 states what the estimator recovers when it is weak. More details can be found in Appendix B.

Target estimands. The model's primary output for product (i) under treatment arm (k) is the conditional counterfactual mean $\mu_k(w, x) := \mathbb{E}[Y(k) \mid W = w, X = x]$, suppressing the unit index i. For a unit i, the model evaluates $\{\mu_k(w_i, x_i)\}_{k=1}^K$ and selects the arm with the lowest expected damage, $\arg\min_k \mu_k(w_i, x_i)$. This decision depends only on relative differences among arms; changing the reporting baseline does not change their ordering. For any ordered pair of arms (k, k'), we define the conditional average treatment effect (CATE):

$$\begin{aligned} \tau_{k,k'}(w, x) &:= \mathbb{E}[Y(k) - Y(k') \mid W = w, X = x] \\ &= \mu_k(w, x) - \mu_{k'}(w, x). \end{aligned} \tag{1}$$

Under assumptions (a)–(c), this quantity is identified. Evaluated for unit i, its CATE is $\tau_{k,k'}(w_i, x_i)$. A common set of per-arm counterfactual means yields transitive contrasts: $\tau_{k,k'}(w, x) + \tau_{k',k''}(w, x) = \tau_{k,k''}(w, x)$. Thus, coherent per-unit ranking requires internally consistent counterfactual levels rather than independently estimated pairwise effects. For a chosen baseline arm b, the pairwise average treatment effect for a focal pair is $\overline{\tau}_{k,b} = \mathbb{E}_{W,X}[\tau_{k,b}(W, X)] = \mathbb{E}[Y(k) - Y(b)]$.

Failure points and MATEL design. Four properties defeat naive causal estimation. First, the outcome is zero-inflated and heavy-tailed (Figure D.4), making Gaussian heads and squared-error loss misspecified. Second, decisions need *absolute* per-product DPMO to price the trade-off and *transitive* contrasts to rank pack types, which standard estimators do not jointly enforce. Third, many pack-type pairs have limited or no overlap, with some products never shipped in certain pack types, a positivity violation under which inverse-propensity and DML estimators diverge (Chernozhukov et al., 2018; Wager & Athey, 2018) (Section 4.1). Fourth, damage depends on high-dimensional, multimodal covariates that require flexible representation learning. MATEL responds with a self-calibrating Tweedie loss, hybrid feature representations, marginal and pairwise targeting for doubly robust calibration, transitive reconciliation into coherent per-arm DPMO levels (Section 3.2), and uncertainty intervals that widen under weak overlap.

2.2 PACKAGING OPTIMIZATION

With the per-arm counterfactuals in hand, pack-type selection becomes a *constrained cost-minimization* problem. Let s index a shipment and $I_s = \{i : \text{product } i \text{ in } s\}$. For product i, let $\mu_{i,j,k}$ be the counterfactual DPMO for outcome $j \in \{\text{Conc}, \text{PS}\}$ under arm k. Let $c_{s,i,j}$ be the corresponding monetary cost per DPMO (scaled by product i's exposure in shipment s), and $h_{s,k}^{\text{mat}}, h_{s,k}^{\text{lab}}, h_{s,k}^{\text{tr}}$

the per-shipment material, labor, and transportation costs. The total cost of shipment s under pack type k is

$$R(s,k) = \sum_{i \in I_s} \sum_{j \in \{\text{Conc},\text{PS}\}} c_{s,i,j}\mu_{i,j,k} + h^{\text{mat}}_{s,k} + h^{\text{lab}}_{s,k} + h^{\text{tr}}_{s,k}. \tag{2}$$

Choosing one pack type per shipment minimizes $\sum_s \sum_{k \in \mathcal{K}_s} a_{s,k} R(s,k)$ over assignments $a_{s,k} \in \{0,1\}$ under three operational constraints: hard *eligibility guardrails* confining each shipment to a feasible set $\mathcal{K}_s$ induced by the products in I_s; per-pack-type *capacity targets*; and a network *throughput limit*, with the latter two coupling shipments through shared resources. Appendix B.1 states the program (9) and its dual decomposition, and gives the contextual multi-armed bandit (arms are the K pack types) of which it is the exploration-free special case.

3 METHODOLOGY

MATEL estimates, for every product and its candidate pack type, the counterfactual DPMOs μ_k and pairwise conditional effects $\tau_{k,b}$ defined in Section 2. We describe the architecture (Section 3.1), the counterfactual estimation and debiasing procedure (Section 3.2), and summarize the inferential guarantees (Section 3.3), where the formal statements and proofs are deferred to Appendix C.

3.1 MATEL ARCHITECTURE

MATEL has the architecture shown in Figure 1. For a product with covariates $z = (w, x)$, a shared encoder produces a representation $\phi = \Phi_\theta(z)$; K per-arm outcome heads map ϕ to counterfactual DPMO Q_k; and a separate propensity encoder maps the confounders w to assignment probabilities g_k. Targeted regularization couples the two through a learnable fluctuation ε. The design generalizes DragonNet in Shi et al. (2019) to the multi-arm, sparse, heavy-tailed setting through three adaptations: K Tweedie heads in place of a binary Gaussian head, a hybrid propensity encoder decoupled from the shared representation, and targeting of pairwise contrasts rather than only marginal means.

Tabular representation backbone. A contribution of MATEL is to treat the nuisance representation as a *pluggable, state-of-the-art tabular encoder*: prior neural causal estimators use generic MLPs Shi et al. (2019); Shalit et al. (2017), whereas packaging damage is driven by $\sim 10^2$ mixed continuous/categorical features with strong nonlinear interactions, where modern tabular architectures match or exceed gradient-boosted trees Gorishniy et al. (2021). Each continuous feature x_j is encoded with a piecewise-linear embedding (PLE) Gorishniy et al. (2022), projected by a learned matrix to a token $\text{PLE}(x_j)$; categorical features use embedding tables. The tokens are concatenated into the backbone input $h^{(0)}$. For the default **ResNet-PLE**, each of n_b residual blocks computes $h^{(\ell+1)} = h^{(\ell)} + \text{Drop}_2\Big(W_2^{(\ell)}\text{Drop}_1\big(\sigma_r(W_1^{(\ell)}\text{BN}(h^{(\ell)}))\big)\Big)$, with σ_r the ReLU and $\phi = h^{(n_b)}$. The **FT-Transformer** alternative appends a `[CLS]` token and applies L multi-head self-attention blocks, taking ϕ as the `[CLS]` embedding. We default to ResNet-PLE: it is cheaper at scale with comparable accuracy (Table D.4), and its smoother, non-adaptive inductive bias is less prone to fitting the factual arm through treatment-correlated shortcuts that inflate the representation-imbalance term of Shalit et al. (2017); Johansson et al. (2016), at parity factual accuracy Gorishniy et al. (2021). The backbone is interchangeable, and both encoders slot into the same causal machinery.

Multimodal augmentation. Each product optionally carries a 768-dim multimodal embedding e_{product} encoded from DINOv2 Oquab et al. (2024) (images) and CLIP ViT-B/32 Radford et al. (2021) (text), used as fixed feature generators. We project it ($p = \text{proj}(e_{\text{product}})$ via $\text{Linear}(768 \to 64)$+ReLU) and fuse $\phi_{\text{fused}} = [\phi; p]$ into the outcome heads, appending the same p to the propensity encoder as an added confounder; 20% embedding dropout keeps predictions robust when it is missing. The augmentation is optional when the tabular attribute features supply required covariates.

Tweedie outcome heads. The shared trunk has K per-arm heads; head k produces a log-mean $f_k(z)$ that a log link maps to counterfactual damage $Q_k(z) = \exp\big(f_k(z)\big)$ on the native DPMO scale. Each head trains only on its factual rows, under the Tweedie negative log-likelihood $\ell_p\big(y, f\big) = -\,y\,\frac{e^{(1-p)f}}{1-p} + \frac{e^{(2-p)f}}{2-p}$, the compound Poisson–Gamma loss suited to the DPMO distribution Jørgensen (1987). We set $p = 1.5$ and initialize each head's output bias to the log of its weighted factual mean DPMO so that sparse heads are stabilized in early epochs.

Hybrid, renormalized propensity encoder. We split $z = (w, x)$ into confounders w (affecting both assignment and damage) and outcome-only features x ($x \perp T \mid w$). The outcome backbone uses both to form ϕ feeding the K heads, while the propensity uses a separate encoder over w alone, without x or ϕ ($\phi \not\to g$): confounders are shared, outcome-only features stay with the outcome. This is the *hybrid* design, versus DragonNet's single shared representation. The propensity emits

$$g_k(w) = \sigma\big(h_k(w)\big), \qquad e_{k,b} = \frac{g_k}{g_k + g_b} = \frac{1}{1+r}, \quad r = \frac{g_b}{g_k}, \tag{3}$$

with $e_{k,b}$ the renormalized pairwise propensity used for targeting. We use independent sigmoids rather than a softmax. Both parameterizations are consistent for the same renormalized pairwise propensity, while one-vs-rest fitting avoids finite-sample simplex coupling between dominant and rare arms (Proposition 1 and Remark 2). Keeping x out of the propensity is deliberate. Prognostic x helps only through the outcome (Theorem 3), while feeding pack-type-varying features into g would break overlap and manufacture positivity violations (Corollary 2 and Table 1).

Clever covariate. On the two-arm overlap subset $\mathcal{I}_{k,b}$, contrasts are targeted through the binary clever covariate

$$H(z, T) = \frac{\mathbf{1}\{T = k\}}{e_{k,b}} - \frac{\mathbf{1}\{T = b\}}{1 - e_{k,b}}, \tag{4}$$

built from the renormalized propensity (3). The same H drives both the pairwise training term below and the inference-time fluctuation of Section 3.2.

Training objective. MATEL is trained with a multi-arm targeted objective that extends targeted regularization Shi et al. (2019) and classical targeted maximum likelihood estimation (TMLE) van der Laan & Rubin (2006) with two elements tailored to damage prediction, a Tweedie outcome loss and a pairwise targeting term:

$$\mathcal{L} = \underbrace{\textstyle\sum_i \omega_i\, \ell_p\big(y_i, f_{T_i}(z_i)\big)}_{\mathcal{L}_Y} + \alpha \underbrace{\textstyle\sum_k \mathrm{BCE}\big(g_k, \mathbf{1}\{T = k\}\big)}_{\mathcal{L}_T} + \beta\, \mathcal{L}_\varepsilon^{\mathrm{marg}} + \gamma\, \mathcal{L}_\varepsilon^{\mathrm{pair}}, \tag{5}$$

with $\alpha = \beta = 1$, $\gamma = \frac{1}{2}$, and shipment weights ω_i. Here $\mathcal{L}_Y$ is the Tweedie loss and $\mathcal{L}_T$ the one-vs-rest propensity loss. Both targeting terms apply the Tweedie loss to fluctuated factual log-means,

$$\mathcal{L}_\varepsilon^{\mathrm{marg}} = \textstyle\sum_i \omega_i\, \ell_p\big(y_i,\ f_{T_i} + \varepsilon_{T_i}/g_{T_i}\big), \quad \mathcal{L}_\varepsilon^{\mathrm{pair}} = \sum_{i \in \mathcal{I}_{k,b}} \omega_i\, \ell_p\big(y_i,\ f_{T_i} + (\varepsilon_k - \varepsilon_b)\, H_i\big). \tag{6}$$

The marginal term generalizes per-treatment targeting to K arms; the pairwise term applies the same clever covariate H (4) during training, over a sampled contrast (k, b) each mini-batch (Remark 12). The debiased counterfactuals Q^* then come from the post-training TMLE fluctuation of Section 3.2.

Guarding against counterfactual collapse. Because each head trains only on its factual rows, arms can learn uniformly low damage and collapse their counterfactuals toward the baseline, flipping sparse-arm ATE signs. Four architectural choices above guard against this: *(i)* we exclude post-treatment operational features (whose distributions are pack-type-determined, e.g. manual workload dominates box packing) from the outcome model, so no head extrapolates from another treatment's data; *(ii)* we aggregate features at the product level so all arms of a unit share one covariate vector and one representation ϕ, and the effect arises solely from head divergence; *(iii)* the hybrid representation (Theorem 3) keeps the propensity from corrupting ϕ and averts the extreme weights that destabilize sparse-arm targeting; and *(iv)* Tweedie heads on the native DPMO scale stop sparse heads from shrinking toward zero, after which the clipped targeting step (7) corrects residual per-arm bias without unbounded inflation on thin overlap (Theorem 1 and Assumption 3).

3.2 Debiased estimation

The plug-in heads Q_k are biased counterfactuals; we debias the contrasts with targeted maximum-likelihood estimation van der Laan & Rubin (2006); van der Laan & Rose (2011) and reconcile the results into a coherent, transitive level output.

The targeting step. Using the clever covariate H (4) on the two-arm overlap subset $\mathcal{I}_{k,b}$, the fluctuation $\hat{\varepsilon}$ solves $\mathbb{P}_{n,kb}\big[\omega H(Y - Q_T^*)\big] = 0$ and yields the targeted counterfactuals and their conditional contrast:

$$Q_k^* = Q_k + \frac{\hat{\varepsilon}}{e_{k,b}}, \qquad Q_b^* = Q_b - \frac{\hat{\varepsilon}}{1 - e_{k,b}}, \qquad \hat{\tau}_{k,b}(w, x) = Q_k^*(w, x) - Q_b^*(w, x). \tag{7}$$

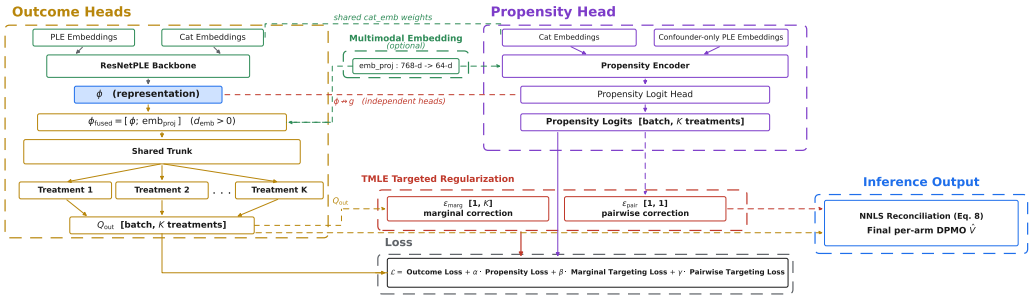


Figure 1: **MATEL architecture.** Hybrid representation with targeted calibration and reconciled output $\hat{V}$.

This is an additive update on the DPMO scale, with $|\hat{\varepsilon}| \leq C_\varepsilon$ (Assumption 3). The per-arm (marginal) estimator is the special case with $H_k = 1/g_k$ on the arm-k rows, giving $Q_k^* = Q_k + \hat{\varepsilon}_k/g_k$ and the per-arm absolute levels.

How the two targeting stages compose. The stages are complementary: per-arm updates calibrate each marginal arm mean, while the pairwise update (7) solves the weighted focal-pair estimating equation. The resulting pairwise calibration is doubly robust under Theorem 2, and differences between targeted counterfactual predictions provide per-unit CATE estimates. The training targeting terms of (5) act as targeted regularization on the pairwise contrast, so the frozen inference-time fluctuation $\hat{\varepsilon}$ that delivers the first-order debiasing is small (Remark 12).

Transitive reconciliation. The debiased contrasts are over-determined ($\binom{K}{2}$ constraints on K levels) and are not mutually transitive. Let $A \in \{-1, 0, 1\}^{\binom{K}{2} \times K}$ be the contrast design ($A_{(k,b),k}=1$, $A_{(k,b),b}=-1$). For fixed confounder values w and feature values x, let $\hat{\tau}$ be the vector of independently debiased pairwise CATE estimates evaluated at (w, x), and let $V \in \mathbb{R}^K_{\geq 0}$ be the per-arm absolute counterfactual levels to be reconciled; AV is then the induced vector of pairwise contrasts, with row (k, b) equal to $V_k - V_b$. For each unit we solve the non-negative least-squares program

$$\hat{V} = \arg\min_{V \geq 0} \ \|AV - \hat{\tau}\|_2^2 + \lambda \|V - \bar{Q}\|_2^2, \tag{8}$$

with $\bar{Q} \in \mathbb{R}^K_{\geq 0}$ the model's marginally targeted per-arm levels and λ a small ridge. The differences of $\hat{V}$ best approximate the debiased pairwise effects (the product CATEs) while the overall level stays near $\bar{Q}$, giving one coherent set of per-pack-type levels for each unit, from which every CATE and pairwise ATE is recovered as a transitive difference (Proposition 4).

3.3 Theoretical Guarantees

We summarize the guarantees supporting MATEL's main methodological contributions; complete statements and proofs are in Appendix C. Let $\rho < \infty$ denote the relative-overlap margin and η the minimum absolute propensity of the two focal arms. Note that ρ can stay small even as $\eta \to 0$.

Stable pairwise learning under sparse overlap. MATEL targets each focal pair using $e_{k,b} = g_k/(g_k + g_b)$. Under relative overlap, $1/\{e_{k,b}(1 - e_{k,b})\} \leq \rho + 2 + \rho^{-1}$, so the targeting weights depend on the balance within the pair rather than the absolute arm propensities. With bounded base predictions and a clipped fluctuation, the resulting corrections and contrasts remain uniformly bounded independently of η (Theorem 1). The pairwise calibration is doubly robust with an $O(1+\rho)$ second-order remainder (Theorem 2). Standard full-population DML retains the raw $1/g_k$ weights; its variance diverges along sparse-arm sequences as $\eta \to 0$, even when the focal-pair share and ρ remain fixed (Lemma 2 and Corollary 3). If clipping binds, MATEL remains bounded and the clipping remainder is controlled (Remark 7 and Proposition 2). The full-population contrast is available from the marginal estimator, which is $\sqrt{n}$-consistent and asymptotically normal under the conditions of Corollary 1 and Proposition 3, with interval width governed by $\mathbb{E}[1/g_k] + \mathbb{E}[1/g_b]$. The pairwise diagnostics in the empirical studies report the subpopulation contrast $\theta_{k,b}$ (13), which averages over units where both arms are attainable. This matches the decision rule: (2) picks one pack type for shipment s from $\mathcal{K}_s$ using product-level estimates $\{\mu_{i,j,k} : i \in I_s\}$ instead of a population average.

Hybrid representation efficiency. When X predicts the outcome but not treatment given W, including it in the outcome model weakly lowers the efficiency bound, while including it in the propensity model provides no gain (Theorem 3 and Corollary 2). A fully shared encoder can also pass strongly treatment-predictive features into the propensity representation and push pairwise propensities toward their boundaries (Remark 10). This motivates MATEL's rich outcome backbone and separate confounder-only propensity encoder; Table 1 measures the finite-sample gain.

Transitive reconciliation and uncertainty. Independently corrected pairwise effects need not correspond to a common set of counterfactual levels. MATEL fits one non-negative, level-anchored vector to all pairwise effects. For $\lambda > 0$ the solution is unique, its differences are transitive, and they induce a coherent ranking of all arms (Proposition 4). The anchor identifies the absolute levels even when the observed pair graph is disconnected. Aggregate ATE uncertainty is reported using product-clustered influence-function standard errors (Proposition 3). For per-unit contrasts, we add a predictive-variance floor under thin overlap (Remark 8). Remark 9 gives the additional conditions for transporting a frozen fluctuation to production.

4 Empirical Studies

MATEL runs in production on AWS SageMaker: training uses `ml.g5.48xlarge` GPU instances, and inference runs Batch Transform over 1000 `ml.c7i.4xlarge` instances scoring full customer orders. We evaluate MATEL on held-out observational data, known-truth synthetic data, and a downstream packaging-decision simulation in a production environment.

4.1 Model Diagnostics on Real-World Data

Input data. MATEL trains on 2.5B shipments after excluding the low-count long tail. Causal diagnostics use a held-out set of 287K products and 755M shipments, restricted to products reliably observed under all pack types. Additional distributional checks over the full scored catalog (11M products and 7B shipments) are reported in Appendix D; they show that factual predictions track observed distributions while raw per-arm rankings retain substantial treatment-selection bias.

The observational reference. For each focal pair (k, b), validation uses only products shipped in both arms. We compare observed DPMO within each product and weight the differences by shipment volume. Comparing within a product removes the pack type selection confounding that reverses the sign of the raw per-arm rates in Table D.3: fragile items are preferentially boxed, so boxes appear the worst. *Directional agreement* means the estimated effect matches this reference's sign; *overlap-qualified* requires sufficient shipment support in both arms.

Observed vs. predicted DPMO distributions. Figures D.2 and D.5 show MATEL's factual predictions align with observed distributions by pack type and product attributes defined in Table B.1. By product attribute, concession DPMO is highest among collectible items, products that contain liquid, and damageable items; PS DPMO is highest for liquids and lowest for magnetic items. Box-A and Box-B have higher observed concession DPMO than paper and poly bags because fragile, high-value products are preferentially assigned to boxes, which is a textbook treatment-selection bias. At the unit level, observed DPMO is zero-inflated and uninformative (Figure D.4), whereas MATEL's mean predictions (Figure D.3) surface the counterfactual effect while capturing the long tail.

Model ablation performances. Table 1 isolates the representation with backbone, features, and data fixed; both MATEL encoders fit the factual outcomes comparably, so the counterfactuals are what separate them. Vanilla DragonNet is a binary net with shared representation, an MSE outcome loss, without uncertainty quantification; DML is tree-based with a linear final-stage. On pairwise ATEs the fully-shared encoder produces propensity-driven weight extremes on sparse arms: for Box-C the renormalized propensity over-separates to $(0.06, 0.91)$, the fluctuation inflates to $|\varepsilon|$=178.1, and effects flip sign or collapse. Decoupling the propensity fixes this: the hybrid encoder moderates the Box-C overlap to $(0.13, 0.21)$, reduces $|\varepsilon|$ from 178.1 to 74.3, eliminates every sign flip, attains $5/5$ directional agreement, and tracks observed effects on well-observed arms (Paper Mailer-C 33.9 vs. 28.8). Baselines degrade under the same sparsity: vanilla DragonNet's binary propensity collapses to the base rate $(0.02, 0.04)$, leaving 3 sign flips and $2/5$ agreement†; DML's inverse-propensity weights diverge on Box-C (-207.8 vs. 3.7), where Box-C is rare relative to Box-A; MATEL's clipped fluctuation and moderated propensities keep the correction bounded. The ordering holds

Table 1: Shared-versus-hybrid encoder ablation with DragonNet and benchmarks on overlap-qualified units. Pairwise ATEs are indexed to the Box-A observed rate = 100; other diagnostic metrics are unscaled.

	MATEL (shared)	MATEL (hybrid)	DragonNet (Vanilla)	DML	Observed
Pairwise ATE vs. Box-A (Conc., indexed)					
Paper Mailer-C	4.5	**33.9**	−15.3	9.7	28.8
Paper Mailer-A	9.2	**18.5**	5.4	7.3	15.1
Paper Mailer-B (Sparse)	−17.1	**35.1**	−8.6	−88.8	18.7
PolyBag-A	25.3	**43.2**	14.1	12.0	82.0
Box-C (Sparse)	0.7	**11.6**	−34.0	−207.8	3.7
CATE vs. Box-A Directional Agreement %					
Paper Mailer-C	60	**87**	47	59	100
Paper Mailer-A	65	**77**	35	49	100
Paper Mailer-B (Sparse)	**35**	34	7	22	100
PolyBag-A	55	**66**	40	52	100
Box-C (Sparse)	32	**39**	7	21	100
CATE vs. Box-A 95% CI Coverage					
Paper Mailer-C	86	**89**	–	78	100
Paper Mailer-A	83	**91**	–	90	100
Paper Mailer-B (Sparse)	89	**97**	–	74	100
PolyBag-A	**96**	95	–	95	100
Box-C (Sparse)	82	**91**	–	67	100
Diagnostics					
ATE direction agrees w/ observed	4/5	**5/5**	2/5	3/5	5/5
Sign flips (vs. observed)	1	**0**	3	2	0
$\lvert\varepsilon\rvert$ (Box-C (Sparse) vs. Box-A)	178.1	**74.3**	2.2†	–	–
g_{binary} (Box-C (Sparse) vs. Box-A)	$(\underline{0.06}, \underline{0.91})$	$(\mathbf{0.13}, \mathbf{0.21})$	$(0.02, 0.04)$	–	–

† Vanilla DN's $\lvert\varepsilon\rvert$ reflects weight-swamp ($1/g$ pinned at the clip, $\approx$100), not stability.

per product: on per-unit CATE directional agreement the hybrid encoder beats DML by $\geq$ 12 pp and vanilla DragonNet by $\geq$ 26 pp on *every* arm (Box-C: 39% vs. 21% and 7%). Its 95% CATE intervals cover 89–97% of product-level observed contrasts across every arm, whereas DML undercovers where overlap is weak (67% on Box-C, 74% on Paper Mailer-B). Across all 45 pairwise contrasts on the 287k-product Singles-concession set, MATEL agrees directionally with the observational reference on 41/45 pairs (91%) versus 28/45 (62%) for DML (Table D.6); MATEL's four disagreements with the observational reference occur on near-zero contrasts, for which sign-based agreement is not informative.

Ground-truth-free effect diagnostics. Following Chernozhukov et al. (2025), we use BLP and GATES to evaluate heterogeneity calibration and a Qini-style damage curve to evaluate ranking; detailed definitions are in Appendix D, and complete results are in Table D.1 and Fig. D.1. MATEL's BLP slopes lie in $[0.79, 1.09]$ with all confidence intervals excluding zero, including on sparse arms, whereas three DML slopes are indistinguishable from zero. MATEL also outperforms random ordering on all nine contrasts, yielding lower realized damage across box-to-bag conversion volumes.

4.2 Validations on Synthetic Data

We construct a five-arm synthetic data-generating process (DGP) with known arm-specific response surfaces, controlled sparse overlap, and heavy-tailed Tweedie noise calibrated to the observed outcome distribution; details are in Appendix E and Table D.7. We compare MATEL with four benchmarks using $\sqrt{\text{PEHE}}$ (Hill, 2011), ATE error ϵ_{ATE}, *Qini* (Radcliffe, 2007), and *Uplift* (Rzepakowski & Jaroszewicz, 2012), where the neural estimators are initialized over five seeds.

MATEL achieves the lowest median ATE error and strongest ranking metrics at every overlap level, recovering 91–92% of oracle uplift with Qini remaining between 0.85 and 0.87. It also has the lowest $\sqrt{\text{PEHE}}$ under weak and severe overlap and is comparable to CausalForest under moderate overlap (1398 ± 166 vs. 1335 ± 143), with its variation across overlap levels within seed spread. Its median ϵ_{ATE} remains between 276 and 353 DPMO across overlap levels, consistent with the bounded behavior in Theorem 1. CausalForest and X-learner trail MATEL on ATE error, *Uplift*, and *Qini* at every overlap level; the X-learner uses Tweedie outcome models and a pseudo-Huber effect stage. Vanilla DragonNet ranks competitively (Qini 0.81–0.83) but misestimates the ATE by 1108–1567 DPMO, which implies that a shared representation with Gaussian head recovers the ordering of the zero-inflated response while its calibration drifts. DML with Tweedie nuisance has the highest PEHE but mixed results for other metrics.

Table 2: **Synthetic validation.** The *minimum focal-pair mass* $q_{\min}$ falls in three levels, at fixed $\rho = 2.30$, $\mathbb{E}[q] = 0.45$, and $n_{k,b} = 26.9$k. Errors are medians over all pairs on the synthetic data before reconciliation. Ranking metrics are for the well-overlapped contrast (Paper Mailer-C vs. Box-A subset, actual ATE 504 DPMO). ± is the IQR half-width over five seeds; uplift is the percentage of oracle uplift recovered.

		all contrasts				*focal contrast*			
		$\sqrt{\text{PEHE}}$ (DPMO)		ϵ_{ATE} (DPMO)		Qini		Uplift (% oracle)	
$q_{\min}$	Method	median	±	median	±	median	±	median	±
10^{-1} (moderate)	**MATEL**	1398	166	**276**	73	**0.87**	0.03	**92**	4
	DragonNet (vanilla)	1942	174	1108	185	0.83	0.01	88	2
	DML	1983	104	624	38	0.62	0.10	54	5
	X-learner	1520	44	509	15	0.66	0.01	67	2
	CausalForest	**1335**	143	454	39	0.81	0.05	81	5
10^{-2} (weak)	**MATEL**	**1301**	184	**330**	72	**0.85**	0.02	**92**	2
	DragonNet (vanilla)	1979	159	1233	120	0.81	0.01	85	4
	DML	2085	87	708	56	0.70	0.02	76	9
	X-learner	1556	23	576	21	0.72	0.01	73	3
	CausalForest	1559	131	481	124	0.78	0.03	82	5
10^{-3} (severe)	**MATEL**	**1257**	193	**353**	43	**0.86**	0.02	**92**	2
	DragonNet (vanilla)	2104	225	1567	184	0.81	0.01	85	1
	DML	2207	65	584	148	0.76	0.06	82	4
	X-learner	1665	35	655	19	0.70	0.01	70	4
	CausalForest	1646	212	493	43	0.80	0.02	84	5

4.3 Simulation Study of MATEL Behaviors in Packaging Optimization

To gauge the production impact of MATEL, we create a production scenario using 10M sampled shipments spanning holiday and off-peak periods. The *baseline* reproduces the rule-based packaging heuristic, while the *treatment* supplies MATEL's estimates to the optimizer in Section 2.2. Both retain eligibility guardrails, with capacity constraints disabled to isolate the estimates' impact. We sweep the damage-protection weights $w \in [0, 10]$ from the default $w = 1$, re-pricing each w-optimized assignment under the fixed w=1 reference objective (a shipment is *harmed* when its decision costs more than baseline at w=1). Table D.2 traces the entitlement vs. w: savings peak at $w = 1$ (3.74 pp) and decline to 1.94 pp at $w = 10$ as the optimizer over-prices damage, while the converted share rises from 19% to 48% but the *improved* fraction plateaus near 25%, so the excess accrues as a growing *harmed* share ($1\% \to 25\%$). At $w = 1$, MATEL yields 26% conversion and a 24.7% net Box-A reduction. We attribute the full entitlement to MATEL because cost-only optimization ($w = 0$) ignores damage and is operationally infeasible.

5 Conclusions

We introduced MATEL, a multi-treatment causal learner for decision problems with heavy-tailed outcomes and sparse overlap. Its hybrid architecture separates rich outcome representations from propensity estimation; pairwise renormalization and targeting provide bounded, doubly robust corrections under relative overlap; and level-anchored reconciliation converts independently corrected effects into coherent per-unit counterfactuals. MATEL therefore supplies a constrained optimizer with both absolute arm outcomes and transitive treatment rankings. On held-out observational data, MATEL agreed with overlap-qualified pairwise directions on 91% of contrasts versus 62% for DML and improved per-unit CATE directional agreement by at least 12 pp. In known-truth synthetic experiments, it recovered 91–92% of oracle uplift, achieved the lowest ATE error at every overlap level, and the lowest PEHE under weak and severe overlap. MATEL is deployed with substantial production impact; because exact figures are confidential, the production-calibrated simulation reports a 3.74 pp network-cost improvement and a 25% reduction in corrugated box usage.

The current system still has two limitations. First, for the sparsest arms, the targeting update may reach its clipping bound, so the resulting effects are bounded extrapolations when overlap is insufficient. Second, the optimizer makes deterministic assignments from predicted means. Uncertainty can reject an unreliable recommendation, but it does not otherwise affect the ranking of feasible actions. A natural next step is a risk-aware online policy that uses the full predictive distribution, balancing expected damage against tail risk and updating decisions as new outcomes arrive.

AI USE STATEMENT

In this work, we used generative AI tools as general-purpose editing and verification aids. Specifically, they were used to improve the readability and concision of author-written text, check grammar, format LaTeX tables, make minor typographical corrections, suggest localized bug fixes in author-written prototype code, flag potential inconsistencies in author-developed mathematical claims, suggest local exposition improvements in the proofs, and locate authoritative references for established theoretical results invoked in those proofs. The authors independently verified all mathematical claims and proofs, tested all suggested code changes, and read the original sources before citing them. We have not used generative AI tools to generate synthetic datasets; develop the theoretical model or conceptual framework; formulate mathematical claims or provide critical ingredients for their proofs; propose or refine hypotheses; design the methodology or experiments; develop or implement the proposed method; select experimental parameters; clean or reformat the data; or interpret results. They were not used to originate scientific content or draft parts of the paper. Tasks involving translation, surveys, interviews, recordings, and qualitative analysis were not applicable. The authors reviewed all AI-assisted edits and take full responsibility for the final content of this paper, including all text, claims, code, results, and artifacts.

ETHICS STATEMENT

This work proposes a causal deep learning framework motivated by packaging assignment in retail fulfillment. It involves no human subjects, personally identifying information, or customer-level data: records are analyzed at the product level, and outcomes are operational defect rates rather than individual attributes. The observational data are proprietary to the industrial partner that operates the fulfillment network and are used with its permission. The framework is deployed in production and delivers substantial, measurable operational and financial impact. To protect confidential business information, Appendix D reports damage rates on an indexed scale and uses anonymized pack-type labels. Similarly, Section 4.3 reports disclosure-safe, production-calibrated simulation results rather than raw deployment figures. We identify two ethical considerations. First, because the optimizer acts on estimated effects, overstating the protective advantage of a cheaper pack type could shift damage onto customers. Eligibility guardrails therefore remain hard constraints (Section 2.2); thin-overlap contrasts are flagged using the uncertainty score in Remark 8; and Table D.2 reports the harmed share alongside the improved share rather than net savings alone. Second, we describe reduced corrugated-box usage as an environmental benefit. The authors declare no competing interests beyond the industrial affiliation disclosed in the non-anonymous version.

REPRODUCIBILITY STATEMENT

The main text and appendix document the method in detail: Section 3.1 specifies the architecture, Tweedie heads, hybrid propensity encoder, and clever covariate; Section 3.2 specifies the targeting step and NNLS reconciliation program; and (5) gives the training objective, with hyperparameters and backbone settings listed in Table D.5. Every formal claim in Section 3.3 is stated with its assumptions and proved in Appendix C, including cases in which the guarantees do not apply (Remark 7 for a binding clip and Remark 9 for a transported fluctuation). Because the empirical studies use proprietary operational data, Section 4.1 presents indexed metrics and Section 4.3 presents production-calibrated simulation estimates. Independent reproduction is enabled through the specified synthetic benchmark in Section 4.2: Appendix E defines the response surfaces, assignment mechanism, overlap sweep, and dispersion calibration; Table E.1 lists all constants; and Algorithm 1 gives the generation procedure. Any compatible covariate table supports end-to-end known-ground-truth evaluation of effect accuracy and ranking under the specified DGP; Table 2 reports the instantiation used in this study. We plan to release the MATEL implementation, including the backbones, Tweedie heads, hybrid propensity encoder, pairwise targeting loss, reconciliation procedure, and synthetic DGP, as open-source code upon publication at ICLR, subject to approval by our institution.

REFERENCES

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019. doi: 10.1214/18-AOS1709.

A. Colin Cameron and Douglas L. Miller. A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372, 2015. doi: 10.3368/jhr.50.2.317.

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. doi: 10.1111/ectj.12097.

Victor Chernozhukov, Mert Demirer, Esther Duflo, and Iván Fernández-Val. Fisher–Schultz Lecture: Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. *Econometrica*, 93(4):1121–1164, 2025. doi: 10.3982/ECTA19303.

Max H. Farrell, Tengyuan Liang, and Sanjog Misra. Deep neural networks for estimation and inference. *Econometrica*, 89(1):181–213, 2021. doi: 10.3982/ECTA16901.

Yury Gorishniy, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. Revisiting deep learning models for tabular data. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, 2021.

Yury Gorishniy, Ivan Rubachev, and Artem Babenko. On embeddings for numerical features in tabular deep learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 2022.

Ben B. Hansen. The prognostic analogue of the propensity score. *Biometrika*, 95(2):481–488, 2008. doi: 10.1093/biomet/asn004.

Jennifer L. Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011. doi: 10.1198/jcgs.2010.08162.

Fredrik D. Johansson, Uri Shalit, and David Sontag. Learning representations for counterfactual inference. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pp. 3020–3029, 2016.

Bent Jørgensen. Exponential dispersion models. *Journal of the Royal Statistical Society: Series B*, 49(2):127–145, 1987.

Jean Kaddour, Aengus Lynch, Qi Liu, Matt J. Kusner, and Ricardo Silva. Causal machine learning: A survey and open problems, 2022. URL `https://arxiv.org/abs/2206.15475`.

Edward H. Kennedy. Semiparametric doubly robust targeted double machine learning: A review. *arXiv preprint arXiv:2203.06469*, 2022.

Sören R. Künzel, Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019. doi: 10.1073/pnas.1804597116.

Abhirup Mondal, Anirban Majumder, and Vineet Chaoji. MEMENTO: Neural model for estimating individual treatment effects for multiple treatments. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. Association for Computing Machinery, 2022.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021. doi: 10.1093/biomet/asaa076.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, et al. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

Nicholas J. Radcliffe. Using control groups to target on predicted lift: Building and assessing uplift models. *Direct Marketing Analytics Journal*, 1:14–21, 2007.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994. doi: 10.1080/01621459.1994.10476818.

Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983. doi: 10.1093/biomet/70.1.41.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. doi: 10.1037/h0037350.

Piotr Rzepakowski and Szymon Jaroszewicz. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32(2):303–327, 2012. doi: 10.1007/s10115-011-0434-0.

Patrick Schwab, Lorenz Linhardt, and Walter Karlen. Perfect match: A simple method for learning representations for counterfactual inference with neural networks. *arXiv preprint arXiv:1810.00656*, 2018. URL `https://arxiv.org/abs/1810.00656`.

Uri Shalit, Fredrik D. Johansson, and David Sontag. Estimating individual treatment effect: Generalization bounds and algorithms. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pp. 3076–3085, 2017.

Claudia Shi, David M. Blei, and Victor Veitch. Adapting neural networks for the estimation of treatment effects. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Anastasios A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, 2006. doi: 10.1007/0-387-37345-4.

Mark J. van der Laan and Sherri Rose. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer, 2011. doi: 10.1007/978-1-4419-9782-1.

Mark J. van der Laan and Daniel Rubin. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1), 2006. doi: 10.2202/1557-4679.1043.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.

Borja Velasco-Regulez and Jesus Cerquides. Hydranet: A neural network for the estimation of multi-valued treatment effects. In *Artificial Intelligence Research and Development*, volume 375 of *Frontiers in Artificial Intelligence and Applications*, pp. 16–27. IOS Press, 2023. doi: 10.3233/FAIA230655.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018. doi: 10.1080/01621459.2017.1319839.

Qin Xiang, Yubai Yuan, Dongyuan Song, Usman J. Wudil, Muktar H. Aliyu, C. William Wester, and Bryan E. Shepherd. Double machine learning to estimate the effects of multiple treatments and their interactions, 2025. URL `https://arxiv.org/abs/2505.12617`.

Zhenyu Zhao and Totte Harinen. Uplift modeling for multiple treatments with cost optimization. In *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 422–431. IEEE, 2019. doi: 10.1109/DSAA.2019.00057.

A BENCHMARK SELECTION

We select benchmarks that cover the principal estimator families relevant to our setting: neural representation learning, orthogonal residualization, meta-learning, and forest-based estimation. Vanilla DragonNet (Shi et al., 2019) serves as the neural representation-learning baseline. It combines a shared representation, outcome and propensity heads, and targeted regularization, and we apply it separately to each focal binary contrast.

For the multi-arm neural comparison, we use shared-layer MATEL as a competitive, architecture-controlled analogue corresponding to the shared-representation design used by multi-arm networks such as HydraNet (Velasco-Regulez & Cerquides, 2023). Shared-layer MATEL jointly learns all K outcome heads while routing the outcome and propensity branches through a common representation. In Table 1, it uses the same data, backbone, Tweedie objective, targeting procedure, and optimization protocol as the hybrid model. The shared vs. hybrid comparison therefore isolates the effect of separating the rich outcome representation from confounder-only propensity estimation while retaining a fully multi-arm neural estimator. Existing multi-arm implementations were not directly transferable to our exposure-weighted Tweedie setting without simultaneously changing the loss, weighting, and training protocol. The architecture-controlled analogue avoids those confounding differences.

DML (Chernozhukov et al., 2018) represents the orthogonal-residualization family, including the residualized-treatment principle underlying the R-learner (Nie & Wager, 2021). The X-learner (Künzel et al., 2019) provides a treatment-effect meta-learner, while causal forest (Athey et al., 2019) provides a nonparametric tree-based comparator. Together, these models provide complementary tests of factual prediction, treatment-effect calibration, effect ranking, and robustness under sparse overlap.

B EXTENDED PROBLEM SETUP

This appendix collects the setup materials referenced from the main text: the pack-type taxonomy (Figure B.1), the product attributes definition (Table B.1), the causal DAG (Figure B.2), and the detailed treatment of the identification assumptions of Section 2.

Table B.1: Product attributes. These are non-mutually-exclusive binary catalog flags; a product may have multiple attributes.

Attribute	Operational definition
Consumable	Intended to be consumed or depleted through normal use.
Collectable	Collectible merchandise for which product or retail-package condition is important.
Granular	Contains loose particles, powder, or granular material.
Human-edible	Food or beverage intended for human consumption.
Magnetic	Contains a magnet or magnetized component.
Sensitive	Requires additional protection because of handling or environmental sensitivity.
Damageable	Classified as particularly susceptible to physical damage during handling or transit.
Liquid	Contains liquid or semi-liquid contents.
Sharp	Contains exposed or packaged sharp edges, points, or components.
Rollable	Can be rolled without damaging the product.
Rectangular	Has an approximately rectangular or cuboid geometry.
Foldable	Can be folded without damaging the product.

Handling the identification assumptions. All three assumptions of Section 2 are non-trivial in our setting and each is addressed explicitly. **SUTVA** is secured for singles by defining treatment and outcomes at the product level, i.e., $I_s = \{i\}$. For multi-item shipments the pack type is a bundle-level treatment shared by all products in I_s; the model conditions on co-shipment composition, while reported product-level potential outcomes are marginalized over its distribution. Each pack-type label denotes a well-defined treatment, and product-level costs are aggregated at the shipment



Figure B.1: Different pack types in the fulfillment network: Boxes, Paper Mailers, PolyBags, and Product Packaging, which differ in material thickness, rigidity, seal, and flexibility.

level in (2). **Unconfoundedness** is targeted by encoding the systematic drivers of pack assignment in confounders W. **Positivity** is the assumption empirically strained for sparse arms. For arms that are operationally prohibited or infeasible under product contexts, positivity is not required. Per-arm minimum thresholds determine which individual contrasts we feed into training and evaluation, and clipping at the arm-pair level bounds the magnitude of the fitted correction. Where support is thin, the model-based uncertainty score of Remark 8 is used as a guardrail on conversion decisions.

B.1 The Assignment Program and its Contextual Bandit Generalization

The constrained program. With $R(s,k)$ the per-shipment cost of (2), pack-type selection over S shipments is the integer program

$$\begin{aligned} \min_{\{a_{s,k}\}} \quad & \sum_{s=1}^{S} \sum_{k \in \mathcal{K}_s} a_{s,k} R(s,k) \\ \text{s.t.} \quad & \sum_{k \in \mathcal{K}_s} a_{s,k} = 1 \qquad \forall s, \qquad a_{s,k} \in \{0,1\}, \\ & \sum_{s=1}^{S} a_{s,k} \le \mathrm{Cap}_k \qquad \forall k, \\ & \sum_{s=1}^{S} \sum_{k \in \mathcal{K}_s} \theta_{s,k} a_{s,k} \le \Theta. \end{aligned} \tag{9}$$

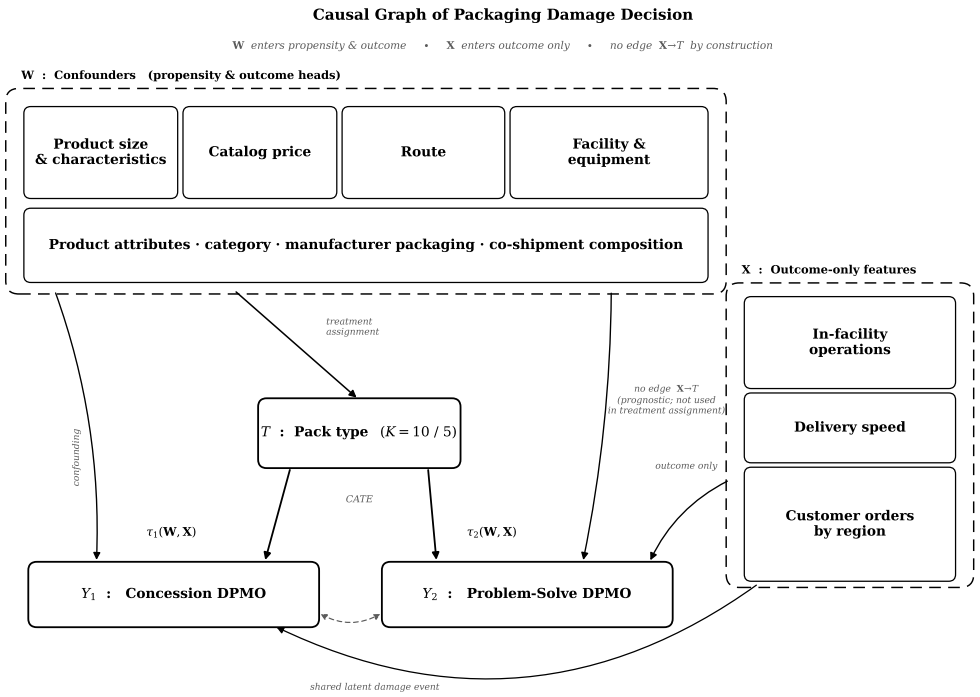


Figure B.2: Causal DAG: confounders W affect both pack-type assignment T and damage Y; feature set X affects only Y.

where $\mathcal{K}_s$ is the feasible set induced by the products in I_s, Cap_k is the per-pack-type capacity target, Θ is the network throughput limit, and $\theta_{s,k}$ is the per-shipment throughput draw. Dualizing the coupling constraints with shadow prices ϕ_k for capacity and ψ for throughput gives a Lagrangian that is separable across shipments. Its relaxation therefore splits into S per-shipment problems,

$$T_s^\star(z_s) = \arg\min_{k \in \mathcal{K}_s} \{R(s,k) + \phi_k + \psi\theta_{s,k}\}.$$

When shipment s is formed, its product-level estimates are aggregated online into $R(s,k)$, and the resulting subproblem is solved using the current shadow prices.

Bandit generalization. The deterministic program of Section 2.2 is the exploration-free special case of a *contextual multi-armed bandit*, the framing that governs pack-type selection once assignments run online with feedback. In that view, the arms are the K pack types, and the context of shipment s combines the product covariates $\{z_i = (w_i, x_i) : i \in I_s\}$ with shipment-level features. Selecting arm k incurs a stochastic cost with mean $R(s,k)$ from (2); a decision rule maps this context to a randomized arm and is judged by cumulative regret against the per-context optimum $\min_{k \in \mathcal{K}_s} R(s,k)$.

Two properties collapse this bandit to the deterministic assignment used in the main text. First, MATEL provides counterfactual estimates for every product under every arm before assignment. When shipment s is formed, the estimates for products in I_s are aggregated online into $R(s,k)$ for every feasible arm. Thus, no feasible arm is left unscored at decision time and there is no trade-off between exploration and exploitation: the regret-minimizing policy is the certainty-equivalent minimizer of (2), subject to the eligibility, capacity, and throughput constraints of (9). Second, each shipment receives the estimated cost-minimizing assignment rather than taking an exploratory action. Capacity and throughput constraints couple shipments through shared shadow prices, and the dual decomposition of Section 2.2 provides a scalable Lagrangian relaxation of the batch.

The framework extends naturally to the genuinely online regime. When damage feedback arrives sequentially and MATEL's counterfactuals carry non-negligible epistemic uncertainty on sparse arms, product-level risk scores $\rho_{i,j,k} = \mathbb{Q}_{1-\alpha}[Y_j(k) \mid z_i]$ can be aggregated over $i \in I_s$ as in (2) to form a risk-averse shipment cost. An optimistic confidence-bound rule could replace the deterministic minimizer by $\arg\min_{k\in\mathcal{K}_s}\{R(s,k) - \kappa\sigma_{s,\cdot,k}\}$, where $\sigma_{s,\cdot,k}$ denotes shipment-level predictive uncertainty. A bootstrap or ensemble approximation could support Thompson sampling over the shipment costs induced by $\{\mu_{i,j,k} : i \in I_s\}$. The uncertainty-dependent coefficient κ could then serve as the exploration schedule.

C Theoretical Analysis: Statements and Proofs

C.1 Setup and assumptions

The observed data are n realizations of $O = (W, X, T, Y)$ with $Z = (W, X)$: confounders W, pre-treatment outcome-only features X with $X \perp T \mid W$ (Hansen, 2008), treatment $T \in \{1, \dots, K\}$, and bounded outcome $Y \in [0, y^{\max}]$ ($y^{\max} \le 10^6$ DPMO). Each row carries a known pre-treatment exposure weight $\omega > 0$ (shipped units or shipments, per Appendix B) and belongs to a product. We assume that ω has a finite second moment, and that no single product dominates the total exposure weight; rows are independent across products and arbitrarily dependent within a product. Arm-conditional regressions are $\mu_k(z) = \mathbb{E}[Y \mid Z{=}z, T{=}k]$ with conditional variances $\sigma_k^2(z) = \mathrm{Var}(Y \mid Z{=}z, T{=}k)$; per-arm propensities are $g_k(w) = \mathbb{P}(T{=}k \mid W{=}w)$ from independent sigmoid heads $g_k = \sigma(f_k)$. The causal targets are the full-population arm means and their contrasts,

$$\theta_k = \mathbb{E}_P[\mu_k(Z)], \qquad \overline{\tau}_{k,b} = \theta_k - \theta_b, \qquad \tau_{k,b}(z) = \mu_k(z) - \mu_b(z), \tag{10}$$

identified from the observed law by conditions (a)–(c) of Section 2. Propensity sufficiency enters only through

$$\mathbb{P}(T{=}k \mid Z) = \mathbb{P}(T{=}k \mid W) = g_k(W), \qquad X \mid (W, T{=}k) \overset{d}{=} X \mid W. \tag{11}$$

For a contrast we fix an arm k and a reference b and work on the focal pair $S_{k,b} = \mathbf{1}\{T \in \{k, b\}\}$, writing $\mathcal{I}_{k,b} = \{i : S_{k,b}[i] = 1\}$, $n_{k,b} = |\mathcal{I}_{k,b}|$ and $\pi_{k,b} = \mathbb{P}(S_{k,b} = 1)$, so $\mathbb{E}[n_{k,b}] = n\pi_{k,b}$, and $P_{k,b}$ for the law of O given $S_{k,b} = 1$. The exposure-weighted focal-pair law $P^{\omega}_{k,b}$ is defined by

$$\mathbb{E}_{P^{\omega}_{k,b}}[h(O)] = \frac{\mathbb{E}_{P_{k,b}}[\omega h(O)]}{\mathbb{E}_{P_{k,b}}[\omega]}.$$

Because ω is pre-treatment, reweighting leaves the conditional outcome regressions and treatment probabilities given Z unchanged. Together with (11), this means that $\mu_t(Z)$ and $e_{k,b}(W)$ are the same under $P_{k,b}$ and $P^{\omega}_{k,b}$. We retain the ordinary empirical operator

$$\mathbb{P}_{n,kb}[h] = \frac{1}{n_{k,b}} \sum_{i \in \mathcal{I}_{k,b}} h(O_i),$$

with the exposure weight ω inside the weighted moments. The renormalized pairwise propensity and the pair odds are

$$e_{k,b}(w) = \frac{g_k(w)}{g_k(w) + g_b(w)}, \qquad r(w) = \frac{g_b(w)}{g_k(w)}, \tag{12}$$

and by (11) and Bayes' rule $e_{k,b}(w) = \mathbb{P}(T{=}k \mid W{=}w, S_{k,b} = 1)$. The subpopulation contrast is

$$\theta_{k,b} = \mathbb{E}_{P^{\omega}_{k,b}}[\mu_k(Z) - \mu_b(Z)], \tag{13}$$

the quantity the pairwise calibration step solves exactly; (10) is the causal target, and the two are distinguished throughout.

Targeted regularization (van der Laan & Rubin, 2006; Shi et al., 2019) fluctuates the outcome heads by a scalar $\widehat{\varepsilon}_{k,b}$ with the renormalized clever covariate

$$H(z, T) = \frac{\mathbf{1}\{T = k\}}{e_{k,b}(w)} - \frac{\mathbf{1}\{T = b\}}{1 - e_{k,b}(w)}, \tag{14}$$

which takes the value $1/e_{k,b} \geq 1$ on arm-k rows and $-1/(1-e_{k,b})$ with modulus at least 1 on arm-b rows, so

$$H^2 \geq 1 \quad \text{on every row of } \mathcal{I}_{k,b}. \tag{15}$$

The fluctuation is fitted on $\mathcal{I}_{k,b}$ by solving

$$\mathbb{P}_{n,kb}[\omega H(Y - Q_T^*)] = 0, \qquad Q_T^* = \widehat{Q}_T + \varepsilon H. \tag{16}$$

Evaluated per arm, this gives the corrected surfaces and contrast

$$\begin{aligned} Q_k^* &= \widehat{Q}_k + \frac{\widehat{\varepsilon}_{k,b}}{e_{k,b}}, \qquad Q_b^* = \widehat{Q}_b - \frac{\widehat{\varepsilon}_{k,b}}{1 - e_{k,b}}, \\ \tau_{k,b}^* &= Q_k^* - Q_b^* = (\widehat{Q}_k - \widehat{Q}_b) + \frac{\widehat{\varepsilon}_{k,b}}{e_{k,b}(1 - e_{k,b})}, \end{aligned} \tag{17}$$

the last equality using $1/e + 1/(1-e) = 1/\{e(1-e)\}$.

Assumption 1 (Bounded outcome)**.** $Y \in [0, y^{\max}]$.

Assumption 2 (Anchored heads)**.** $\widehat{Q}_t(z) \in [0, y^{\max}]$ *for all* z *and* t*. The Tweedie log link ensures non-negative predictions, and the upper bound is enforced by output clipping.*

Assumption 3 (Clipped fluctuation)**.** *The implemented fluctuation is* $\widehat{\varepsilon}_{k,b} = \mathrm{clip}(\widehat{\varepsilon}_{k,b}^{\mathrm{unc}}, -C_\varepsilon, C_\varepsilon)$ *for a fixed* $C_\varepsilon < \infty$*, where* $\widehat{\varepsilon}_{k,b}^{\mathrm{unc}}$ *solves* (16) *without constraint.*

Assumption 4 (Relative overlap)**.** (i) *The true pair odds* $r_0(w) = g_{0,b}(w)/g_{0,k}(w)$ *satisfy* $r_0(w) \in [\rho^{-1}, \rho]$ *for a finite* $\rho \geq 1$ *on the target support of* W*.* (ii) *The fitted odds* $\hat{r}(w) = \hat{g}_b(w)/\hat{g}_k(w)$ *are confined to the same interval by a prespecified pairwise propensity bound, which is asymptotically inactive under (i).*

Part (i) is a condition on the data-generating process; part (ii) is an algorithmic device. For the comparison with raw-propensity weighting we write

$$\eta = \operatorname*{ess\,inf}_w \min_{t \in \{k,b\}} g_t(w) \tag{18}$$

for the *absolute* overlap margin and $\pi_k = \mathbb{P}(T{=}k) = \mathbb{E}[g_k(W)]$ for the marginal share of arm k. Assumption 4 can hold with $\rho = O(1)$ while $\eta \to 0$: two arms may be simultaneously rare in absolute terms yet comparably rare relative to each other.

Remark 1 (Relative overlap and pair support). Assumption 4(i) forces $g_{0,b}(w) > 0$ wherever $g_{0,k}(w) > 0$ and conversely, so it does not weaken pairwise positivity, condition (c) of Section 2. It is weaker only than *uniform* positivity across all K arms. What it buys is that the correction bound of Lemma 1 does not degrade as the focal pair's shared mass $\pi_{k,b}$ shrinks.

C.2 Estimator construction

Two estimators of $\overline{\tau}_{k,b}$ appear below and are never conflated. The *marginal* estimator is the augmented inverse-propensity form of Robins et al. (1994), using the raw per-arm propensities on the full sample,

$$\begin{aligned} \widehat{\overline{\tau}}_{k,b}^{\mathrm{marg}} = \mathbb{P}_n \Big[&\widehat{Q}_k - \widehat{Q}_b \\ &+ \frac{\mathbf{1}\{T=k\}}{\hat{g}_k(W)}(Y - \widehat{Q}_k) - \frac{\mathbf{1}\{T=b\}}{\hat{g}_b(W)}(Y - \widehat{Q}_b) \Big], \end{aligned} \tag{19}$$

with nuisances $(\widehat{Q}_t, \hat{g}_t)$ fitted on a training split and applied to held-out rows, each product assigned entirely to one split. The *pairwise-corrected* plug-in averages the corrected surfaces (17) over the target sample, $\widehat{\overline{\tau}}_{k,b}^{\mathrm{corr}} = \mathbb{P}_n^{\mathrm{target}}[\tau_{k,b}^*(W,X)]$, where $\mathbb{P}_n^{\mathrm{target}}$ is the weighted empirical average over all rows of the target sample, not only those in $\mathcal{I}_{k,b}$. The influence-function results attach to (19).

For the marginal target define the arm-k and contrast influence functions

$$D_k(O; \mu, g) = \mu_k(Z) - \theta_k + \frac{\mathbf{1}\{T=k\}}{g_k(W)}\{Y - \mu_k(Z)\}, \tag{20}$$

$$D_{k,b}(O; \mu, g) = D_k(O; \mu, g) - D_b(O; \mu, g), \tag{21}$$

and for the exposure-weighted subpopulation contrast (13), under $P^{\omega}_{k,b}$,

$$D^*(O;\mu,e) = \big(\mu_k(Z) - \mu_b(Z) - \theta_{k,b}\big) + \frac{\mathbf{1}\{T=k\}}{e_{k,b}(W)}\big(Y - \mu_k(Z)\big) - \frac{\mathbf{1}\{T=b\}}{1-e_{k,b}(W)}\big(Y - \mu_b(Z)\big). \tag{22}$$

Every appearance of D^* below refers to (22).

Under (11) with $g_k, g_b > 0$ almost surely, $D_{k,b}$ has mean zero and variance

$$\mathrm{Var}(D_{k,b}) = \mathrm{Var}(\mu_k(Z) - \mu_b(Z)) + \mathbb{E}\left[\frac{\sigma_k^2(Z)}{g_k(W)}\right] + \mathbb{E}\left[\frac{\sigma_b^2(Z)}{g_b(W)}\right], \tag{23}$$

the standard independent-row influence-function variance of the augmented estimator (Robins et al., 1994; Tsiatis, 2006): the weighted residuals $R_t = \mathbf{1}\{T=t\}g_t^{-1}(W)\{Y - \mu_t(Z)\}$ satisfy $\mathbb{E}[R_t \mid Z] = 0$ with $\mathbb{E}[R_t^2] = \mathbb{E}[\sigma_t^2(Z)/g_t(W)]$, are mutually uncorrelated because no row carries both arms, and are uncorrelated with $\mu_k - \mu_b$.

The targeting equation (16) is affine in ε with slope $-\mathbb{P}_{n,kb}[\omega H^2]$. Since $\mathbb{P}_{n,kb}[\omega H^2] \geq \mathbb{P}_{n,kb}[\omega] > 0$, it has the unique solution

$$\widehat{\varepsilon}^{\mathrm{unc}}_{k,b} = \frac{\mathbb{P}_{n,kb}\Big[\omega H(Y - \widehat{Q}_T)\Big]}{\mathbb{P}_{n,kb}[\omega H^2]}. \tag{24}$$

which vanishes in the population when $\widehat{Q}_t = \mu_t$ for $t \in \{k, b\}$.

Proposition 1 (Propensity parameterization: invariance and sparse-arm decoupling)**.** *Suppose $\hat{g}_k(W), \hat{g}_b(W)$ are consistent for $g_k(W), g_b(W)$ with $g_k + g_b > 0$ on the target support. Then*

$$\hat{e}_{k,b}(W) = \frac{\hat{g}_k(W)}{\hat{g}_k(W) + \hat{g}_b(W)} \xrightarrow{p} \frac{g_k(W)}{g_k(W) + g_b(W)} = \mathbb{P}\big(T{=}k \mid W, S_{k,b} = 1\big), \tag{25}$$

whether the raw probabilities are produced by a calibrated joint softmax or by calibrated one-vs-rest heads.

Proof. By Bayes' rule and (11),

$$\begin{aligned}\mathbb{P}(T{=}k \mid W, S_{k,b} = 1) &= \frac{\mathbb{P}(T{=}k, S_{k,b} = 1 \mid W)}{\mathbb{P}(S_{k,b} = 1 \mid W)} \\ &= \frac{\mathbb{P}(T{=}k \mid W)}{\mathbb{P}(T{=}k \mid W) + \mathbb{P}(T{=}b \mid W)} \\ &= \frac{g_k}{g_k + g_b},\end{aligned} \tag{26}$$

where the middle step uses $\{T{=}k\} \subset \{S_{k,b} = 1\}$ and the disjointness of $\{T{=}k\}$ and $\{T{=}b\}$. The map $h(u,v) = u/(u+v)$ is continuous on $\{u, v \geq 0,\, u + v > 0\}$, so the continuous mapping theorem transfers convergence in probability of $(\hat{g}_k, \hat{g}_b)$ to $\hat{e}_{k,b}$. □

Remark 2 (Softmax versus independent sigmoids). Proposition 1 is asymptotic and does not imply that the two parameterizations agree in finite samples. For matched logits they do not: with $f_k = 2$, $f_b = 0$, one-vs-rest heads give $\sigma(2)/\{\sigma(2) + \sigma(0)\} = 0.638$ whereas a softmax gives $\sigma(f_k - f_b) = \sigma(2) = 0.881$. We use one-vs-rest heads because each logit is fit without the simplex coupling that ties a rare arm's score to the dominant arms, and we report propensity calibration and rare-arm weight diagnostics for both parameterizations.

C.3 Statements for overlap, boundedness, and double robustness

Lemma 1 (Renormalization identity)**.** *On $\{g_k + g_b > 0\}$ the pairwise propensity depends on the two arms only through their ratio,*

$$e_{k,b} = \frac{1}{1+r}, \qquad 1 - e_{k,b} = \frac{r}{1+r}, \tag{27}$$

and is invariant to their common scale: $(g_k, g_b) \mapsto (\lambda g_k, \lambda g_b)$ *leaves* $e_{k,b}$ *unchanged for every* $\lambda > 0$*. Under Assumption 4,*

$$\begin{aligned} \frac{1}{e_{k,b}} &= 1 + r \le 1 + \rho, \qquad \frac{1}{1 - e_{k,b}} = 1 + \frac{1}{r} \le 1 + \rho, \\ \frac{1}{e_{k,b}(1 - e_{k,b})} &= r + 2 + \frac{1}{r} \le \left(\sqrt{\rho} + \tfrac{1}{\sqrt{\rho}}\right)^2, \end{aligned} \tag{28}$$

and none of these bounds involves $\min_t g_t$ *or any absolute overlap margin.*

Proof. Dividing numerator and denominator of (12) by $g_k > 0$ gives $e_{k,b} = 1/(1 + r)$, whence $1 - e_{k,b} = 1 - 1/(1 + r) = r/(1 + r)$. Scale invariance is immediate because r is a ratio. Taking reciprocals, $1/e_{k,b} = 1 + r$ and $1/(1 - e_{k,b}) = (1 + r)/r = 1 + 1/r$, and both are bounded by $1 + \rho$ on $r \in [\rho^{-1}, \rho]$. For the product,

$$\begin{aligned} e_{k,b}(1 - e_{k,b}) &= \frac{1}{1 + r} \cdot \frac{r}{1 + r} = \frac{r}{(1 + r)^2}, \\ \frac{1}{e_{k,b}(1 - e_{k,b})} &= \frac{(1 + r)^2}{r} = \frac{1 + 2r + r^2}{r} = r + 2 + \frac{1}{r}. \end{aligned} \tag{29}$$

Let $\phi(r) = r + 1/r$ on $(0, \infty)$. Then $\phi'(r) = 1 - r^{-2}$, which vanishes only at $r = 1$, and $\phi''(r) = 2r^{-3} > 0$, so ϕ is strictly convex. A strictly convex function on a compact interval attains its maximum at an endpoint, and $\phi(\rho) = \phi(\rho^{-1}) = \rho + \rho^{-1}$. Hence $\phi(r) \le \rho + \rho^{-1}$ on $[\rho^{-1}, \rho]$ and $1/\{e_{k,b}(1 - e_{k,b})\} \le \rho + 2 + \rho^{-1}$. Finally $(\sqrt{\rho} + 1/\sqrt{\rho})^2 = \rho + 2 + \rho^{-1}$, giving the stated form. No step referenced g_k or g_b except through r. □

Lemma 2 (Inverse-propensity divergence)**.** *Suppose* (11) *holds and* $\sigma_k^2(Z) \ge \underline{v} > 0$ *on the support of arm* k*. Then*

$$\mathrm{Var}(R_k) = \mathbb{E}\left[\frac{\sigma_k^2(Z)}{g_k(W)}\right] \;\ge\; \underline{v}\,\mathbb{E}\left[\frac{1}{g_k(W)}\right] \;\ge\; \frac{\underline{v}}{\pi_k}, \tag{30}$$

and consequently, by (23)*, the semiparametric efficiency bound (Tsiatis, 2006) for* $\overline{\tau}_{k,b}$ *at sample size* n *is at least* $n^{-1}\underline{v}\,\mathbb{E}[1/g_k(W)]$*. Then we have:* (i) *If* $\pi_k \to 0$ *the bound diverges at rate* $1/\pi_k$*.* (ii) *For fixed* $r > 0$ *and every* $0 < \pi_k < 1/(1 + r)$*,* $\mathbb{E}[1/g_k]$ *is not controlled by* π_k*: for every* $M < \infty$ *there is a law with marginal arm share* π_k *and constant pair odds* $g_b/g_k = r$ *for which* $\mathbb{E}[1/g_k] > M$*.*

Proof. The first equality is the arm-k term of (23). The first inequality replaces σ_k^2 by its lower bound $\underline{v}$, valid because $1/g_k > 0$. For the second, the map $u \mapsto 1/u$ is convex on $(0, \infty)$, so Jensen's inequality gives $\mathbb{E}[1/g_k(W)] \ge 1/\mathbb{E}[g_k(W)]$, and $\mathbb{E}[g_k(W)] = \mathbb{E}[\mathbb{P}(T{=}k \mid W)] = \mathbb{P}(T{=}k) = \pi_k$ by the tower property. The efficiency-bound statement follows from (23), since its remaining terms are non-negative. Claim (i) is immediate from (30).

For (ii), fix $r > 0$ and $0 < \pi_k < 1/(1 + r)$. Let $\alpha = \{1 - (1 + r)\pi_k\}/2$ and let $B \sim \mathrm{Bernoulli}(\alpha)$ be a component of W. For sufficiently small $\delta > 0$, define

$$\begin{aligned} v_\delta &= \frac{\pi_k - \alpha\delta}{1 - \alpha}, \qquad g_k(W) = \delta B + v_\delta(1 - B), \\ g_b(W) &= r\, g_k(W), \qquad g_u(W) = 1 - (1 + r) g_k(W). \end{aligned} \tag{31}$$

Because $v_\delta < \pi_k/(1 - \alpha) < 1/(1 + r)$, these are valid probabilities for sufficiently small δ. Moreover, $\mathbb{E}[g_k] = \alpha\delta + (1{-}\alpha)v_\delta = \pi_k$, while $g_b/g_k = r$, so $e_{k,b} = 1/(1{+}r)$ and $\rho = \max(r, r^{-1})$ remain fixed. Also, $\pi_{k,b} = (1 + r)\pi_k$ remains fixed, and hence so does the distribution of $n_{k,b}$. Finally,

$$\mathbb{E}\left[\frac{1}{g_k(W)}\right] = \frac{\alpha}{\delta} + \frac{1 - \alpha}{v_\delta} \;\xrightarrow[\delta \downarrow 0]{}\; \infty.$$

Thus $\mathbb{E}[1/g_k]$ can exceed any prescribed M while π_k, ρ, $e_{k,b}$, and the distribution of $n_{k,b}$ remain fixed. □

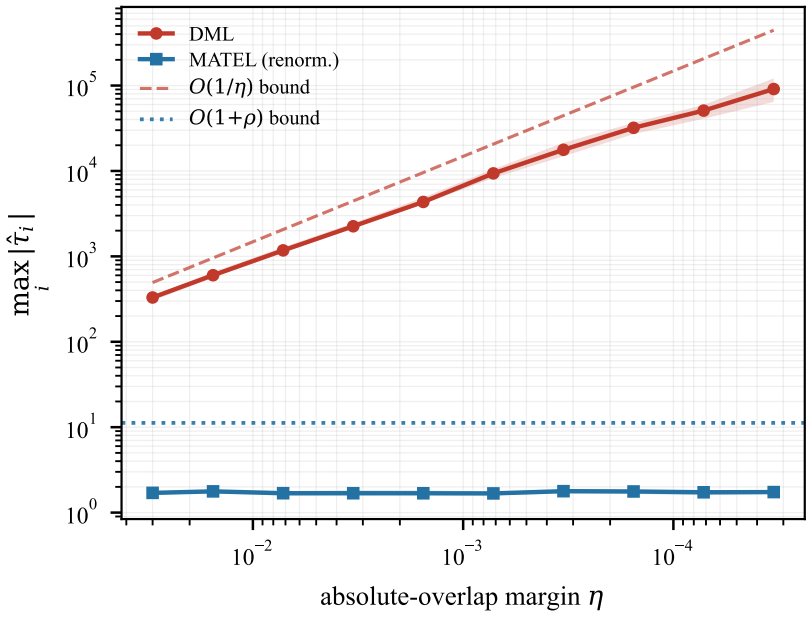


Figure C.1: **Synthetic illustration of the overlap dependence.** Three synthetic arms: the focal pair plus a dominant unrelated arm with $g_k(w) = q(w)\, e_{k,b}(w)$, $g_b(w) = q(w)\{1 - e_{k,b}(w)\}$, and the remaining mass $1 - q(w)$ on the third arm. This separates the two controls: $\mathbb{E}[q(W)]$ sets the focal-pair count $n_{k,b}$, while $e_{k,b}$ sets the relative margin ρ. We hold both fixed ($n_{k,b} \approx 14$k, $\rho \approx 2.3$) and shrink q on a segment of 30% of units, driving the absolute margin η of (18) down three decades. This is the one axis on which the two estimators' variances do not move together. Both estimators receive the correct propensities. DML's largest per-unit estimate $\max_i |\hat{\tau}_i|$ diverges along the $O(1/\eta)$ bound (dashed), while the renormalized, ε-clipped MATEL estimator stays flat under the $O(1+\rho)$ bound (dotted). Both bounds are predicted from the data-generating process, not fitted. Medians are over 25 seeds, bands are IQRs; illustrates Lemma 2 and Theorem 1.

Theorem 1 (Uniform boundedness)**.** *Under Assumptions 1 to 4, for $t \in \{k, b\}$,*

$$\begin{aligned} \left|Q_t^* - \widehat{Q}_t\right| &\le C_\varepsilon(1+\rho), \qquad Q_t^* \in \left[-C_\varepsilon(1+\rho),\ y^{\max} + C_\varepsilon(1+\rho)\right], \\ \left|\tau_{k,b}^*\right| &\le y^{\max} + C_\varepsilon\left(\sqrt{\rho} + \tfrac{1}{\sqrt{\rho}}\right)^2, \end{aligned} \tag{32}$$

the targeting correction is at most $C_\varepsilon(1+\rho)$ per arm, so the targeted counterfactuals and their contrast lie in an envelope determined only by C_ε, ρ, and $y_{\max}$. The bounds are uniform in the arm shares π_k, π_b and the absolute overlap margin η; they depend only on how balanced the two arms are relative to each other, as measured by ρ.

Proof. From (17), $|Q_k^* - \widehat{Q}_k| = |\widehat{\varepsilon}_{k,b}|/e_{k,b}$. Bounding the numerator by Assumption 3 and the reciprocal by Lemma 1 gives $|Q_k^* - \widehat{Q}_k| \le C_\varepsilon(1+\rho)$; the arm-$b$ case is identical using $1/(1-e_{k,b}) \le 1+\rho$. Combining with $\widehat{Q}_t \in [0, y^{\max}]$ (Assumption 2) gives the displayed interval, where the lower endpoint is negative because the additive update is not sign-constrained. For the contrast, the third identity in (17) gives

$$\left|\tau_{k,b}^*\right| \le \left|\widehat{Q}_k - \widehat{Q}_b\right| + \frac{|\widehat{\varepsilon}_{k,b}|}{e_{k,b}(1-e_{k,b})} \le y^{\max} + C_\varepsilon\left(\sqrt{\rho} + \tfrac{1}{\sqrt{\rho}}\right)^2, \tag{33}$$

using $|\widehat{Q}_k - \widehat{Q}_b| \le y^{\max}$ from Assumption 2 and the third bound of (28). Throughout, the propensities g_k and g_b appeared only through their ratio $r = g_b/g_k$; shrinking both by a common factor leaves every bound unchanged, thus, the result holds as $\eta \to 0$. □

Remark 3 (Boundedness is deterministic). Theorem 1 is an arithmetic consequence of the clip and Assumption 4; it holds for any $\widehat{Q}$ satisfying Assumption 2. It is the per-unit guarantee a downstream optimizer consumes.

Remark 4 (Which sample size). The fluctuation $\widehat{\varepsilon}_{k,b}$ is estimated only from the focal-pair observations, so its precision is governed by $n_{k,b}$ rather than the full sample size n. For the sparse contrasts in Table D.1, the weighted effective sample size is only a few percent of that of the best-supported contrast; using n would therefore give a misleading impression of precision. The table reports relative ESS, computed as $\mathrm{ESS}_{k,b} = \frac{\left(\sum_{i\in\mathcal{I}_{k,b}} \omega_i\right)^2}{\sum_{i\in\mathcal{I}_{k,b}} \omega_i^2}$.

The next theorem is the double-robustness property of the pairwise calibration; see Kennedy (2022) for a survey of the broader semiparametric estimators.

Theorem 2 (Bounded double robustness)**.** *Let $\bar{e}, \bar{\mu}_t$ be the limiting fitted nuisances and $e_0, \mu_{0,t}$ the truth, and let D^* be as in* (22)*. Under $P^{\omega}_{k,b}$,*

$$\mathbb{E}_0\big[D^*(O;\bar{\mu},\bar{e})\big] = R, \quad R = \mathbb{E}_0\left[(\bar{e}-e_0)\left(\frac{\bar{\mu}_k-\mu_{0,k}}{\bar{e}} + \frac{\bar{\mu}_b-\mu_{0,b}}{1-\bar{e}}\right)\right]. \tag{34}$$

Hence (i) *$R = 0$ if either $\bar{e} = e_0$ or $\bar{\mu}_t = \mu_{0,t}$ for $t \in \{k, b\}$; and* (ii) *under Assumption 4, with all norms taken in $L_2(P^{\omega}_{k,b})$,*

$$|R| \le (1+\rho)\,\|\bar{e}-e_0\|_2\Big(\|\bar{\mu}_k-\mu_{0,k}\|_2 + \|\bar{\mu}_b-\mu_{0,b}\|_2\Big), \tag{35}$$

which is finite irrespective of π_k and π_b. The corresponding raw-propensity remainder carries $1/\bar{g}_k$ and $1/\bar{g}_b$ in place of $1/\bar{e}$ and $1/(1-\bar{e})$, and admits no bound in terms of ρ alone.

Proof. Step 1 (expand the two weighted residual terms). Condition on Z and $S_{k,b} = 1$, then on T:

$$\mathbb{E}_0\left[\frac{\mathbf{1}\{T=k\}}{\bar{e}}(Y-\bar{\mu}_k)\right] = \mathbb{E}_0\left[\frac{\mathbb{P}_0(T{=}k \mid Z, S_{k,b}=1)}{\bar{e}}\,\mathbb{E}_0\big[Y-\bar{\mu}_k \mid Z, T{=}k\big]\right] \tag{36}$$

$$= \mathbb{E}_0\left[\frac{e_0}{\bar{e}}\big(\mu_{0,k}-\bar{\mu}_k\big)\right], \tag{37}$$

$$\mathbb{E}_0\left[\frac{\mathbf{1}\{T=b\}}{1-\bar{e}}(Y-\bar{\mu}_b)\right] = \mathbb{E}_0\left[\frac{1-e_0}{1-\bar{e}}\big(\mu_{0,b}-\bar{\mu}_b\big)\right], \tag{38}$$

where in each case $\bar{e}$ is W-measurable and passes outside the inner expectation, the conditional treatment probabilities are e_0 and $1-e_0$ by (11) and Bayes' rule, and $\mathbb{E}_0[Y \mid Z, T{=}t] = \mu_{0,t}$ by definition.

Step 2 (assemble). Using (22), (37), (38) and $\theta_{k,b} = \mathbb{E}_0[\mu_{0,k} - \mu_{0,b}]$,

$$\begin{aligned}\mathbb{E}_0[D^*(\bar{\mu},\bar{e})] &= \mathbb{E}_0[\bar{\mu}_k-\bar{\mu}_b] - \mathbb{E}_0[\mu_{0,k}-\mu_{0,b}] + \mathbb{E}_0\left[\frac{e_0}{\bar{e}}(\mu_{0,k}-\bar{\mu}_k)\right] - \mathbb{E}_0\left[\frac{1-e_0}{1-\bar{e}}(\mu_{0,b}-\bar{\mu}_b)\right] \\ &= \mathbb{E}_0\left[(\bar{\mu}_k-\mu_{0,k})\Big(1-\frac{e_0}{\bar{e}}\Big)\right] - \mathbb{E}_0\left[(\bar{\mu}_b-\mu_{0,b})\Big(1-\frac{1-e_0}{1-\bar{e}}\Big)\right].\end{aligned} \tag{39}$$

Step 3 (identify the common factor). Elementary algebra gives

$$\begin{aligned}1-\frac{e_0}{\bar{e}} &= \frac{\bar{e}-e_0}{\bar{e}}, \\ 1-\frac{1-e_0}{1-\bar{e}} &= \frac{(1-\bar{e})-(1-e_0)}{1-\bar{e}} = -\frac{\bar{e}-e_0}{1-\bar{e}},\end{aligned} \tag{40}$$

so the second term changes sign and both terms carry the factor $(\bar{e}-e_0)$:

$$\begin{aligned}\mathbb{E}_0[D^*(\bar{\mu},\bar{e})] = \mathbb{E}_0&\left[(\bar{e}-e_0)\frac{\bar{\mu}_k-\mu_{0,k}}{\bar{e}}\right] \\ &+ \mathbb{E}_0\left[(\bar{e}-e_0)\frac{\bar{\mu}_b-\mu_{0,b}}{1-\bar{e}}\right],\end{aligned} \tag{41}$$

which is (34). The remainder is a sum of products of two nuisance errors, hence second order.

Step 4 (double robustness). If $\bar{e} = e_0$ the factor $(\bar{e} - e_0)$ vanishes pointwise; if $\bar{\mu}_t = \mu_{0,t}$ for both arms then both parenthesised differences vanish. Either way $R = 0$, giving (i).

Step 5 (bound). By the triangle inequality applied to (34) and then the Cauchy–Schwarz inequality on each of the two resulting terms,

$$\begin{aligned} |R| \le & \left\| \frac{\bar{e} - e_0}{\bar{e}} \right\|_2 \|\bar{\mu}_k - \mu_{0,k}\|_2 \\ & + \left\| \frac{\bar{e} - e_0}{1 - \bar{e}} \right\|_2 \|\bar{\mu}_b - \mu_{0,b}\|_2. \end{aligned} \tag{42}$$

By Lemma 1 applied to $\bar{e}$ (legitimate under Assumption 4(ii)), $1/\bar{e} \le 1 + \rho$ and $1/(1 - \bar{e}) \le 1 + \rho$ almost surely, so both leading norms are at most $(1+\rho)\|\bar{e} - e_0\|_2$, giving (35). For the raw-propensity comparison the same computation with g in place of e produces $\|(\bar{g}_k - g_{0,k})/\bar{g}_k\|_2$, and $1/\bar{g}_k$ admits no bound in terms of ρ; Lemma 2(ii) exhibits laws with ρ fixed and $\mathbb{E}[1/g_k]$ arbitrarily large. □

Remark 5 (Relation to the full-population estimator). Theorem 2 is a statement about the exposure-weighted focal-pair target $\theta_{k,b}$ under $P^{\omega}_{k,b}$: the pairwise step solves the estimating equation of the *subpopulation* contrast (13) with a second-order remainder bounded by the relative margin. The reported pairwise-corrected estimator $\widehat{\overline{\tau}}^{\,\mathrm{corr}}_{k,b}$ averages the corrected conditional surfaces over the target covariate distribution, and Proposition 2 bounds its deviation. Full-population influence-function inference is supplied by Proposition 3 for the marginal estimator (19).

Proposition 2 (Bounded bias)**.** *Define*

$$\begin{aligned} b_Q(z) &= \big|(\bar{\mu}_k - \bar{\mu}_b)(z) - \tau_{k,b}(z)\big|, \\ \Delta_{n,k,b}(z) &= \big|(\widehat{Q}_k - \widehat{Q}_b)(z) - (\bar{\mu}_k - \bar{\mu}_b)(z)\big|, \\ a_{n,k,b} &= \big(|\widehat{\varepsilon}^{\mathrm{unc}}_{k,b}| - C_\varepsilon\big)_+. \end{aligned}$$

Under Assumptions 1 to 4, for each fixed z in the target support,

$$\big|\tau^*_{k,b}(z) - \tau_{k,b}(z)\big| \le b_Q(z) + \Delta_{n,k,b}(z) + C_\varepsilon\Big(\sqrt{\rho} + \tfrac{1}{\sqrt{\rho}}\Big)^2. \tag{43}$$

Moreover,

$$\begin{aligned} \big|\tau^{*\mathrm{clip}}_{k,b}(z) - \tau^{*\mathrm{unc}}_{k,b}(z)\big| &\le a_{n,k,b}\Big(\sqrt{\rho} + \tfrac{1}{\sqrt{\rho}}\Big)^2, \\ \big|\widehat{\overline{\tau}}^{\,\mathrm{corr,clip}}_{k,b} - \widehat{\overline{\tau}}^{\,\mathrm{corr,unc}}_{k,b}\big| &\le a_{n,k,b}\Big(\sqrt{\rho} + \tfrac{1}{\sqrt{\rho}}\Big)^2. \end{aligned} \tag{44}$$

Proof. Subtracting $\tau_{k,b}(z)$ from (17) and adding and subtracting $(\bar{\mu}_k - \bar{\mu}_b)(z)$ gives

$$\begin{aligned} \tau^*_{k,b}(z) - \tau_{k,b}(z) = & \big\{(\bar{\mu}_k - \bar{\mu}_b)(z) - \tau_{k,b}(z)\big\} \\ & + \big\{(\widehat{Q}_k - \widehat{Q}_b)(z) - (\bar{\mu}_k - \bar{\mu}_b)(z)\big\} \\ & + \frac{\widehat{\varepsilon}_{k,b}}{e_{k,b}(w)\{1 - e_{k,b}(w)\}}. \end{aligned}$$

The triangle inequality identifies the first two terms as $b_Q(z)$ and $\Delta_{n,k,b}(z)$. By Assumption 3 and Lemma 1, the final term satisfies

$$\left| \frac{\widehat{\varepsilon}_{k,b}}{e_{k,b}(w)\{1 - e_{k,b}(w)\}} \right| \le C_\varepsilon\Big(\sqrt{\rho} + \tfrac{1}{\sqrt{\rho}}\Big)^2,$$

which proves (43).

For the clipping result, the clip operator satisfies

$$\big|\mathrm{clip}(u, -C, C) - u\big| = (|u| - C)_+,$$

so

$$\big|\widehat{\varepsilon}_{k,b} - \widehat{\varepsilon}^{\mathrm{unc}}_{k,b}\big| = a_{n,k,b}.$$

The clipped and unclipped contrasts differ only through this coefficient. Therefore, by (17) and Lemma 1,

$$\left|\tau^{*\mathrm{clip}}_{k,b}(z) - \tau^{*\mathrm{unc}}_{k,b}(z)\right| = \frac{\left|\widehat{\varepsilon}_{k,b} - \widehat{\varepsilon}^{\mathrm{unc}}_{k,b}\right|}{e_{k,b}(w)\{1-e_{k,b}(w)\}} \le a_{n,k,b}\Big(\sqrt{\rho} + \tfrac{1}{\sqrt{\rho}}\Big)^2.$$

The bound is uniform over the target support, so averaging it over the target distribution gives the second inequality in (44). □

Remark 6 (Interpretation of the bound). The decomposition separates the limiting outcome-model bias $b_Q(z)$ from the finite-sample estimation error $\Delta_{n,k,b}(z)$. If $\widehat{Q}_t(z) \to_P \bar{\mu}_t(z)$ for $t \in \{k, b\}$, then $\Delta_{n,k,b}(z) = o_P(1)$, leaving $b_Q(z)$ and the bounded targeting adjustment in (43).

Remark 7 (Interior vs. binding clip). Theorem 2 presumes an interior fluctuation, so that (16) holds exactly and the calibration is EIF-orthogonal for $\theta_{k,b}$. When the clip binds, the fitted coefficient differs from its unconstrained value by $a_{n,k,b}$, so the moment equation is no longer solved exactly and the resulting contrast remainder is bounded by (44). Clipping contributes an additional $O_P(a_{n,k,b})$ remainder, while the estimator remains uniformly bounded by Theorem 1 rather than diverging.

We next establish asymptotic linearity for the full-population marginal estimator (19).

Corollary 1 (Asymptotic linearity of the marginal estimator)**.** *Assume the nuisances are fitted on a training split disjoint from the rows entering the average, $\hat{g}_t \ge \underline{g} > 0$, $\mathrm{Var}(D_{k,b}) < \infty$, and the product-rate condition*

$$\sum_{t\in\{k,b\}} \|\hat{g}_t - g_{0,t}\|_2 \, \|\widehat{Q}_t - \mu_{0,t}\|_2 = o_P(n^{-1/2}), \tag{45}$$

which deep nuisance estimators attain under standard smoothness conditions (Farrell et al., 2021). Then $\widehat{\overline{\tau}}^{\mathrm{marg}}_{k,b}$ is asymptotically linear,

$$\widehat{\overline{\tau}}^{\mathrm{marg}}_{k,b} - \overline{\tau}_{k,b} = \mathbb{P}_n\big[D_{k,b}(O;\mu_0,g_0)\big] + o_P(n^{-1/2}), \tag{46}$$

and for independent rows, $\sqrt{n}(\widehat{\overline{\tau}}^{\mathrm{marg}}_{k,b} - \overline{\tau}_{k,b}) \rightsquigarrow \mathcal{N}(0, \mathrm{Var}(D_{k,b}))$, with $\mathrm{Var}(D_{k,b})$ given by (23)*.*

This is the standard sample-splitting argument for augmented estimators (Chernozhukov et al., 2018, Thm. 3.1): the von Mises expansion (van der Vaart, 1998) of $\widehat{\overline{\tau}}^{\mathrm{marg}}_{k,b} - \overline{\tau}_{k,b}$ has a second-order remainder controlled by (45) after Cauchy–Schwarz and an empirical-process term that vanishes because the nuisances are independent of the rows entering the average. Splitting on whole products makes that independence hold; for within-product dependence, Proposition 3 applies the product-level central limit theorem with the cluster-sandwich estimator (47).

Proposition 3 (Confidence intervals for the reported contrasts)**.** *Let $\mathcal{C}$ index products, $L = |\mathcal{C}|$, and let $\hat{D}_{k,b,i}$ be the plug-in influence values. Assume* (46)*, that products are independent with $L \to \infty$, that $\mathbb{E}[(\sum_{i\in c} D_{k,b,i})^2]$ is uniformly bounded over c, that $\max_c\big(\sum_{i\in c}|D_{k,b,i}|\big)^2 \big/ \sum_c\big(\sum_{i\in c}|D_{k,b,i}|\big)^2 \to_P 0$, and that $\hat{D}_{k,b} \to D_{k,b}(\mu_0,g_0)$ in L_2. Then*

$$\widehat{\mathrm{Var}}\big(\widehat{\overline{\tau}}^{\mathrm{marg}}_{k,b}\big) = \frac{1}{n^2}\sum_{c\in\mathcal{C}}\Big(\sum_{i\in c}\hat{D}_{k,b,i}\Big)^2 \tag{47}$$

is the cluster-robust sandwich estimator at the product level (Cameron & Miller, 2015); it is consistent for $\mathrm{Var}(\widehat{\overline{\tau}}^{\mathrm{marg}}_{k,b})$, and the associated Wald intervals have asymptotically nominal coverage. Under Assumption 4 alone, finite interval width is not guaranteed because the independent-row variance (23) *can diverge with $\mathbb{E}[1/g_k] + \mathbb{E}[1/g_b]$ (Lemma 2).*

Consistency of (47) and nominal Wald coverage are the standard cluster-robust results (Cameron & Miller, 2015) applied to the influence values at the product level, using (46) to replace the estimator by its linear expansion and the stated moment and negligibility conditions to invoke the product-level central limit theorem.

Remark 8 (Model uncertainty on thin overlap). Proposition 3 covers the aggregate estimator $\widehat{\overline{\tau}}^{\,\mathrm{marg}}_{k,b}$ only. For fixed-(w,x) conditional contrasts we report a model-based uncertainty score combining the influence-function variance with a predictive-variance floor $\mathrm{se} = \sqrt{\mathrm{se}_\tau^2 + \mathrm{se}_Q^2}$. That score is calibrated empirically rather than derived and is used for ranking and guardrail recommendations.

Remark 9 (Transported targeting). In production the fluctuation is frozen at its development value and applied as part of the prediction rule. Corollary 1 and Proposition 3 concern an evaluation estimator whose targeting equation uses held-out factual outcomes; they do not by themselves provide doubly robust inference or interval coverage after transporting a frozen fluctuation to a new covariate distribution, where an additional first-order term $\mathbb{E}_{\mathrm{new}}[H(\mu_0 - Q_T^*)]$ appears, bounded by Theorem 1. Transported predictions are therefore evaluated as model outputs, distribution shift and support are monitored operationally, and the fluctuation is refitted on a schedule.

C.4 Representation design

Theorem 3 (Hybrid-representation efficiency)**.** *Assume* (11) *and* $g_k, g_b > 0$. *For a covariate set* $S \in \{W, (W,X)\}$, *let* V_S *be the asymptotic variance of the one-step estimator* (19) *built from* S*-measurable outcome regressions and a* W*-measurable propensity, so that by* (23) *applied with* Z *replaced by* S,

$$V_S = \mathrm{Var}\big(\mu_k(S) - \mu_b(S)\big) + \mathbb{E}\left[\frac{\sigma_k^2(S)}{g_k(W)}\right] + \mathbb{E}\left[\frac{\sigma_b^2(S)}{g_b(W)}\right]. \tag{48}$$

Under (11), $V_{W,X}$ *coincides with the variance of the efficient influence function for* $\overline{\tau}_{k,b}$ *in the nonparametric model (Tsiatis, 2006), while* V_W *is the variance obtained when only the confounders are used in the outcome model; the comparison below is therefore between two architectures, not between two statistical models. Then* (i) *the propensity is unchanged,* $g_t(W,X) = g_t(W)$*; and* (ii)

$$\begin{aligned} V_W - V_{W,X} = \mathbb{E}\Big[& \tfrac{1-g_k}{g_k}\,\mathrm{Var}\big(\mu_k(W,X) \mid W\big) \\ & + \tfrac{1-g_b}{g_b}\,\mathrm{Var}\big(\mu_b(W,X) \mid W\big) \\ & + 2\,\mathrm{Cov}\big(\mu_k(W,X), \mu_b(W,X) \mid W\big)\Big] \;\geq\; 0, \end{aligned} \tag{49}$$

with no sign condition on the covariance. Prognostic X *therefore weakly lowers the bound through the outcome model while adding nothing to the propensity.*

Proof. (i) is (11).

Step 1 (within-arm law of total variance). Fix $t \in \{k, b\}$. Applying the law of total variance to Y given $(W, T{=}t)$, decomposing over X,

$$\begin{aligned} \sigma_t^2(W) = \mathrm{Var}(Y \mid W, T{=}t) = \mathbb{E}\big[\mathrm{Var}(Y \mid W, X, T{=}t) \mid W, T{=}t\big] \\ + \mathrm{Var}\big(\mathbb{E}[Y \mid W, X, T{=}t] \mid W, T{=}t\big). \end{aligned} \tag{50}$$

By (11) the conditional law of X given $(W, T{=}t)$ equals that given W, so both outer moments may be taken with respect to $X \mid W$:

$$\sigma_t^2(W) = \mathbb{E}\big[\sigma_t^2(W,X) \mid W\big] + \mathrm{Var}\big(\mu_t(W,X) \mid W\big). \tag{51}$$

This is the only place propensity sufficiency is used, and the identity fails without it.

Step 2 (weighted noise terms). Since g_t is W-measurable, dividing (51) by $g_t(W)$ and taking expectations gives, by the tower property,

$$\mathbb{E}\left[\frac{\sigma_t^2(W)}{g_t}\right] - \mathbb{E}\left[\frac{\sigma_t^2(W,X)}{g_t}\right] = \mathbb{E}\left[\frac{\mathrm{Var}(\mu_t(W,X) \mid W)}{g_t}\right]. \tag{52}$$

Step 3 (heterogeneity term). By (11) and the tower property, $\mathbb{E}[\mu_t(W,X) \mid W] = \mu_t(W)$. Hence the law of total variance applied to $\mu_k(W,X) - \mu_b(W,X)$ gives

$$\mathrm{Var}\big(\mu_k(W,X) - \mu_b(W,X)\big) = \mathrm{Var}\big(\mu_k(W) - \mu_b(W)\big) + \mathbb{E}\big[\mathrm{Var}(\mu_k - \mu_b \mid W)\big], \tag{53}$$

so that $\mathrm{Var}(\mu_k(W) - \mu_b(W)) - \mathrm{Var}(\mu_k(W,X) - \mu_b(W,X)) = -\mathbb{E}[\mathrm{Var}(\mu_k - \mu_b \mid W)]$.

Step 4 (combine). Substituting Steps 2 and 3 into (48),

$$V_W - V_{W,X} = -\mathbb{E}\big[\mathrm{Var}(\mu_k - \mu_b \mid W)\big] + \mathbb{E}\left[\frac{\mathrm{Var}(\mu_k \mid W)}{g_k}\right] + \mathbb{E}\left[\frac{\mathrm{Var}(\mu_b \mid W)}{g_b}\right], \tag{54}$$

and expanding $\mathrm{Var}(\mu_k - \mu_b \mid W) = \mathrm{Var}(\mu_k \mid W) + \mathrm{Var}(\mu_b \mid W) - 2\,\mathrm{Cov}(\mu_k, \mu_b \mid W)$ and collecting the two $\mathrm{Var}(\mu_t \mid W)$ terms as $\big(g_t^{-1} - 1\big)\,\mathrm{Var}(\mu_t \mid W) = \frac{1-g_t}{g_t}\,\mathrm{Var}(\mu_t \mid W)$ yields the identity in (49).

Step 5 (non-negativity, unconditionally). Put $U_t = \mu_t(W, X) - \mathbb{E}[\mu_t(W, X) \mid W]$, so that $\mathrm{Var}(U_t \mid W) = \mathrm{Var}(\mu_t \mid W)$ and $\mathrm{Cov}(U_k, U_b \mid W) = \mathrm{Cov}(\mu_k, \mu_b \mid W)$, and set $a = \frac{1-g_k}{g_k} > 0$, $b = \frac{1-g_b}{g_b} > 0$. Because the arms are mutually exclusive, $g_k + g_b \le 1$, hence

$$\begin{aligned} ab - 1 &= \frac{(1-g_k)(1-g_b) - g_k g_b}{g_k g_b} \\ &= \frac{1 - g_k - g_b + g_k g_b - g_k g_b}{g_k g_b} = \frac{1 - g_k - g_b}{g_k g_b} \;\ge\; 0. \end{aligned} \tag{55}$$

Now complete the square: for any $a > 0$,

$$\mathrm{Var}\Big(\sqrt{a}\,U_k + \tfrac{1}{\sqrt{a}}U_b \;\Big|\; W\Big) = a\,\mathrm{Var}(U_k \mid W) + \tfrac{1}{a}\,\mathrm{Var}(U_b \mid W) + 2\,\mathrm{Cov}(U_k, U_b \mid W), \tag{56}$$

so the conditional integrand of (49) equals

$$\begin{aligned} &a\,\mathrm{Var}(U_k \mid W) + b\,\mathrm{Var}(U_b \mid W) + 2\,\mathrm{Cov}(U_k, U_b \mid W) \\ &\quad = \mathrm{Var}\Big(\sqrt{a}\,U_k + \tfrac{1}{\sqrt{a}}U_b \;\Big|\; W\Big) + \Big(b - \tfrac{1}{a}\Big)\,\mathrm{Var}(U_b \mid W). \end{aligned} \tag{57}$$

The first term is a variance, hence non-negative; the second is non-negative because $b - 1/a = (ab-1)/a \ge 0$ by (55). Taking expectations over W gives $V_W - V_{W,X} \ge 0$ with no assumption on the sign of the covariance. When $K = 2$ we have $g_k + g_b = 1$, so $ab = 1$, the second term vanishes, and (49) reduces to the familiar binary-treatment identity. □

Corollary 2 (Refinement of DragonNet's fully-shared representation Shi et al. (2019))**.** *For prognostic X satisfying* (11)*, the variance reduction from modelling all covariates is realised entirely through the outcome regressions (Theorem 3(ii)), while by (i) the propensity gains nothing from X. A fully shared representation that routes X into the propensity head therefore cannot lower the bound. This* motivates *a rich outcome representation with a parsimonious propensity encoder. Attaining $V_{W,X}$ additionally requires the nuisance-rate condition* (45)*, so the shared-versus-hybrid comparison is settled empirically in Table 1.*

Remark 10 (Why decoupling is necessary). Theorem 3 presumes (11). *If* a fully shared representation admits into the fitted propensity components that are strongly treatment-predictive at the aggregation grain, *then* the fitted $\hat{e}_{k,b}$ can be driven toward $\{0, 1\}$ even where the arms genuinely overlap, in which case $1/\{\hat{e}_{k,b}(1 - \hat{e}_{k,b})\}$ inflates and the remainder bound (35) degrades through ρ. We state this as the mechanism our architecture is designed to preclude and test it in Table 1.

C.5 Transitivity

Proposition 4 (Transitive reconciliation of contrasts)**.** *Fix confounder values w and feature values x. Let $\mathcal{O} \subseteq \{(k, k') : k > k'\}$ be the available pairs with contrast operator $A \in \{-1, 0, 1\}^{|\mathcal{O}| \times K}$, $(Av)_p = v_{k_p} - v_{k'_p}$; let $\hat{\tau}(w, x)$ collect the debiased pairwise CATE estimates evaluated at the same (w, x); and let $\bar{Q}(w, x) \in \mathbb{R}^K_{\ge 0}$ be the marginally targeted per-arm levels. For $\lambda > 0$ define*

$$\begin{aligned} \hat{V}(w, x) &= \arg\min_{v \ge 0}\; f(v), \\ f(v) &= \|Av - \hat{\tau}(w, x)\|_2^2 + \lambda \|v - \bar{Q}(w, x)\|_2^2, \end{aligned} \tag{58}$$

and $\hat{\tau}^{\mathrm{rec}}_{k,k'} = \hat{V}_k - \hat{V}_{k'}$. Then: (i) *the minimizer exists and is unique;* (ii) *$\hat{\tau}^{\mathrm{rec}}_{k,k'} + \hat{\tau}^{\mathrm{rec}}_{k',k''} = \hat{\tau}^{\mathrm{rec}}_{k,k''}$ for every triple, so the induced relation is a coherent weak ordering of the arms, with ties permitted;* (iii) *with $\widetilde{A} = [A; \sqrt{\lambda} I]$ and $\widetilde{y} = [\hat{\tau}; \sqrt{\lambda}\bar{Q}]$ we have $f(v) = \|\widetilde{A}v - \widetilde{y}\|_2^2$, and the unconstrained minimizer*

$v_0 = (A^\top A + \lambda I)^{-1}(A^\top \hat{\tau} + \lambda \bar{Q})$ *satisfies that* $\widetilde{A} v_0$ *is the Euclidean orthogonal projection of* $\widetilde{y}$ *onto* $\mathrm{range}(\widetilde{A})$*;* (iv) $\mathrm{rank}(A) = K - c$ *where* c *is the number of connected components of the graph on* $\{1, \dots, K\}$ *with edge set* $\mathcal{O}$*, so the anchor term is* required*, not optional, to determine the level within each component; and* (v) *if the graph is connected, the constraint is inactive, and* $\lambda \downarrow 0$*, then* $A\hat{V}$ *converges to the ordinary least-squares projection of* $\hat{\tau}$ *onto* $\mathrm{range}(A)$.

Proof. (i) The Hessian is $\nabla^2 f(v) = 2(A^\top A + \lambda I)$, and for $u \neq 0$

$$u^\top (A^\top A + \lambda I) u = \|Au\|_2^2 + \lambda \|u\|_2^2 \geq \lambda \|u\|_2^2 > 0, \tag{59}$$

so f is strictly convex. It is also coercive, since $f(v) \geq \lambda \|v - \bar{Q}\|_2^2 \to \infty$ as $\|v\| \to \infty$. The feasible set $\{v \geq 0\}$ is non-empty, closed and convex, so a minimizer exists by coercivity and lower semicontinuity, and is unique by strict convexity.

(ii) For any $v \in \mathbb{R}^K$ the differences telescope: $(v_k - v_{k'}) + (v_{k'} - v_{k''}) = v_k - v_{k''}$. Because every reported contrast is a difference of coordinates of the single vector $\hat{V}$, the identity holds for every triple irrespective of the data and irrespective of whether $\hat{\tau}$ is itself coherent. Defining $k \succeq k'$ iff $\hat{V}_k \leq \hat{V}_{k'}$ inherits reflexivity, transitivity and totality from the order on $\mathbb{R}$; it is a total preorder rather than a total order because $\hat{V}_k = \hat{V}_{k'}$ is possible, and indeed likely when the non-negativity constraint is active at zero on several sparse arms.

(iii) Stacking, $\|\widetilde{A}v - \widetilde{y}\|_2^2 = \|Av - \hat{\tau}\|_2^2 + \lambda \|v - \bar{Q}\|_2^2 = f(v)$. Dropping $v \geq 0$, the normal equations of this ordinary least-squares problem are $\widetilde{A}^\top \widetilde{A} v = \widetilde{A}^\top \widetilde{y}$, i.e. $(A^\top A + \lambda I) v = A^\top \hat{\tau} + \lambda \bar{Q}$, whose matrix is positive definite by the display in (i) and hence invertible, giving v_0. That $\widetilde{A} v_0$ is the orthogonal projection of $\widetilde{y}$ onto $\mathrm{range}(\widetilde{A})$ is the defining property of the least-squares solution. Note the projection is of the *augmented* vector $\widetilde{y}$.

(iv) $Av = 0$ holds iff $v_k = v_{k'}$ for every $(k, k') \in \mathcal{O}$, i.e. iff v is constant on each connected component of the pair graph. The space of such v has dimension c, so $\mathrm{rank}(A) = K - \dim \ker A = K - c$. In particular $A^\top A$ is singular for every $c \geq 1$, so at $\lambda = 0$ the objective does not determine the level within a component; the anchor $\lambda \|v - \bar{Q}\|^2$ supplies it, and for $c \geq 2$ all cross-component comparisons are determined by $\bar{Q}$ and λ rather than by any pairwise evidence. This case must be reported when minimum-count filtering disconnects $\mathcal{O}$.

(v) If the graph is connected then $\ker A = \mathrm{span}\{\mathbf{1}\}$ and A has full column rank on $\mathbf{1}^\perp$. Restricting to that subspace, v_0 depends continuously on λ at $\lambda = 0^+$, so $Av_0 \to \Pi_{\mathrm{range}(A)} \hat{\tau}$; if the constraint is inactive then $\hat{V} = v_0$ for all small λ, giving the claim. All contrasts are identified in the limit even though the additive level is not. □

Remark 11 (Weighting the reconciliation). Equation (58) weights all pairs equally, so a contrast estimated from a few focal-pair rows influences the reconstructed levels as strongly as one estimated from many. Because the natural objective is the generalized least-squares criterion $(Av - \hat{\tau})^\top \widehat{\Sigma}^{-1} (Av - \hat{\tau})$ with $\widehat{\Sigma}$ the covariance of $\hat{\tau}$, and because pairs sharing an arm are correlated so that Σ is not diagonal, the unweighted version should be read as a computational choice.

Remark 12 (The pairwise training loss). The training objective (5) contains a pairwise targeted-regularization term $\mathcal{L}_\varepsilon^{\mathrm{pair}}$ that fluctuates the Tweedie log-mean, and its stationarity condition differs from the inference-time moment (16). Since $\partial_f \ell_p(y, f) = -y e^{(1-p)f} + e^{(2-p)f} = e^{(1-p)f}(e^f - y)$, differentiating $\mathcal{L}_\varepsilon^{\mathrm{pair}}$ at the unfluctuated point gives

$$\sum_{i \in \mathcal{I}_{k,b}} \omega_i \, H_i \, Q_i^{1-p} \big(Q_i - Y_i\big) = 0, \qquad p = \tfrac{3}{2} \implies \text{weight } Q_i^{-1/2}, \tag{60}$$

whereas (16) carries no Q-dependent weight; the two coincide when Q is constant across focal-pair rows. This separation is deliberate: Tweedie deviance provides a distribution-matched training regularizer, whereas the post-training exposure-weighted native-DPMO equation targets the influence-function score for $\theta_{k,b}$. The fitted $\widehat{\varepsilon}_{k,b}$ denotes this inference-time fluctuation, to which orthogonality and double robustness apply; reconciliation enforces nonnegative final levels and transitivity by construction.

C.6 What pairwise renormalization removes

Let $q_{k,b}(w) = g_k(w) + g_b(w)$ denote the probability that a unit receives either focal treatment. Then $g_k = q_{k,b}e_{k,b}$, $g_b = q_{k,b}(1 - e_{k,b})$, and therefore

$$\frac{1}{g_k} + \frac{1}{g_b} = \frac{1}{q_{k,b}} \left\{ \frac{1}{e_{k,b}} + \frac{1}{1 - e_{k,b}} \right\} = \frac{1}{q_{k,b}e_{k,b}(1 - e_{k,b})}. \tag{61}$$

Raw marginal weighting is thus sensitive to two distinct sources of instability: how rarely the focal pair is available, through $q_{k,b}$, and how imbalanced the two arms are within that pair, through $e_{k,b}$. Pairwise renormalization conditions on membership in the focal pair and removes the first factor from the correction weights.

Corollary 3 (Separation of absolute and relative overlap)**.** *Fix a focal pair and let* $\pi_{k,b} = \mathbb{P}(S_{k,b} = 1) = \mathbb{E}[q_{k,b}(W)]$*, so that* $\mathbb{E}[n_{k,b}] = n\pi_{k,b}$*. Suppose* $\sigma_t^2(Z) \geq \underline{v} > 0$ *for* $t \in \{k, b\}$*.*

(i) Raw marginal weighting. *The full-population efficiency bound for* $\overline{\tau}_{k,b}$ *is at least*

$$\frac{\underline{v}}{n}\mathbb{E}\left[\frac{1}{q_{k,b}(W)e_{k,b}(W)\{1 - e_{k,b}(W)\}}\right].$$

Consequently, controlling the focal-pair share and relative arm balance does not control the raw bound when $q_{k,b}(W)$ *can approach zero.*

(ii) Pairwise-renormalized weighting. *Under Assumption 4,*

$$\sup_w \frac{1}{e_{k,b}(w)\{1 - e_{k,b}(w)\}} \leq \rho + 2 + \rho^{-1}.$$

This bound depends only on the relative balance of the two focal arms and not on $q_{k,b}(w)$ *or* $\mathbb{E}[1/g_k(W)]$*.*

(iii) What renormalization does not remove. *The witness family in* (31) *holds* $\pi_{k,b}$*,* $e_{k,b}$*,* ρ*, and the distribution of* $n_{k,b}$ *fixed while the raw bound in part (i) diverges. The bound in part (ii) remains fixed. Renormalization nevertheless cannot compensate for limited focal-pair support, summarized by* $n_{k,b}$ *and weighted ESS (Remark 4).*

Proof. Equation (61), together with Lemma 2, gives part (i). Part (ii) follows from Lemma 1. In the witness family (31), $q_{k,b}(W) = (1 + r)g_k(W)$ and $e_{k,b} = 1/(1 + r)$ is constant. As $\delta \downarrow 0$, $\mathbb{E}[1/q_{k,b}(W)] = (1 + r)^{-1}\mathbb{E}[1/g_k(W)] \to \infty$, while $\pi_{k,b}$, $e_{k,b}$, ρ, and the distribution of $n_{k,b}$ remain fixed. Note that at $\rho = 2.30$, the relative-weight bound is $2.30 + 2 + 1/2.30 \approx 4.73$. This proves the separation in part (iii). □

Remark 13 (Scope and matched baselines). This result isolates the weighting mechanism; it does not equate the pairwise correction with full-population influence-function inference. A pairwise DML or TMLE estimator fit on $\mathcal{I}_{k,b}$ using $e_{k,b}$ shares the bound in part (ii). The pairwise construction also retains the outcome-model dependence quantified by b_Q in Proposition 2.

D Complementary Empirical Results

This appendix collects the results referenced from Sections 4.1 and 4.2: the shared-vs-hybrid encoder ablation (Table 1), the singles counterfactual-DPMO table (Table D.3), the factual-prediction distribution figures (Figures D.2 to D.5), the doubly robust curve (Figure D.1), ground-truth-free metrics BLP slope and GATES (Table D.1), and the backbone comparison (Table D.4).

Reducing plastic usage in network. Many retailers are transitioning their packaging suite from plastic to paper bags (Figure B.1). We apply MATEL to estimate the concession and PS implications of moving from PolyBag or Paper Mailer-A to Paper Mailer-B (Table D.3; all rates indexed to the Box-A observed rate = 100). On *problem solve*, Paper Mailer-B is on par with PolyBag-A and both materially outperform Paper Mailer-A. On *concession*, Paper Mailer-A and Paper Mailer-B are comparable and both slightly beat PolyBag-A. The empirical evidence supports the environmentally motivated initiative to transition from plastic to paper bags.

Ground-truth-free diagnostics for causal effect ranking. For each focal contrast (k, b) and each $i \in \mathcal{I}_{k,b}$, let $Z_i = (W_i, X_i)$ and define the pair-specific AIPW pseudo-outcome

$$\psi_i^{k,b} = Q_k^*(Z_i) - Q_b^*(Z_i) + \frac{\mathbf{1}\{T_i = k\}}{e_{k,b}(W_i)}\{Y_i - Q_k^*(Z_i)\} - \frac{\mathbf{1}\{T_i = b\}}{1 - e_{k,b}(W_i)}\{Y_i - Q_b^*(Z_i)\}.$$

Its exposure-weighted mean over $\mathcal{I}_{k,b}$ provides the doubly robust focal-pair reference used by the Best Linear Predictor (BLP) and Sorted Group ATE (GATES) diagnostics (Chernozhukov et al., 2025; Zhao & Harinen, 2019). These diagnostics evaluate heterogeneity calibration and targeting quality. Table D.1 reports both metrics against DML. MATEL's BLP slopes lie in $[0.79, 1.09]$, with all confidence intervals excluding zero, including for the sparse Box-C and Paper Mailer-B contrasts. By comparison, the DML slopes are indistinguishable from zero for Box-B, Box-C, and Paper Mailer-B, and overshoot on Paper Mailer-C and PolyBag-A. MATEL's damage-deferral coefficient exceeds DML's on all nine contrasts (PolyBag-B: 3.24 vs. 1.70), so its ordering incurs less realized DPMO across conversion volumes from more protective boxes to less protective flexible mailers (Figure D.1a).

The BLP slope and GATES measure two aspects of causal validity: *(1) Calibration of heterogeneity:* following Chernozhukov et al. (2025), the Best Linear Predictor (BLP) regresses $\psi^{k,b}$ on the demeaned $\hat{\tau}_{k,b}(W, X)$ (slope near 1 with CI excluding 0 = calibrated heterogeneity), and Sorted Group ATE (GATES) bins units by $\hat{\tau}$ into a monotone staircase along the identity (Figure D.1b), confirming the ranking is real. *(2) Targeting quality:* knowing that Box-A is most protective, every substitution raises damage and the task is to route each conversion to the least-harmful pack; a Qini-style cost curve Zhao & Harinen (2019) sorts shipments by predicted damage and accumulates the realized DPMOs. Staying *below* random (Figure D.1a) means less realized DPMO at every volume, summarized by a damage-deferral coefficient (area saved over random; higher is better). We report effective sample size (ESS) as a support diagnostic.

Backbone prediction accuracy. Table D.4 compares backbones on factual-prediction accuracy at three shipment-volume thresholds (tabular features only; MAE *improvement* over a shipment-weighted per-product average, as absolute rates are confidential). Improvement falls as the threshold rises—a headroom effect, since the baseline already fits high-volume cells well—so the largest gains accrue on the under-shipped minority pack types MATEL targets for conversion. ResNet and FT-Transformer hold their advantage on high-volume cells (30.1% and 28.9% at >10,000) while XGBoost erodes fastest (24.4%). FT-Transformer is better in two columns; ResNet wins at the highest-volume threshold and trains faster at only half the cost (6h vs. 13h), motivating it as the default; XGBoost is a factual reference only, since as a tree it cannot host the shared representation the propensity and Tweedie heads require. Hyperparameters are in Table D.5.

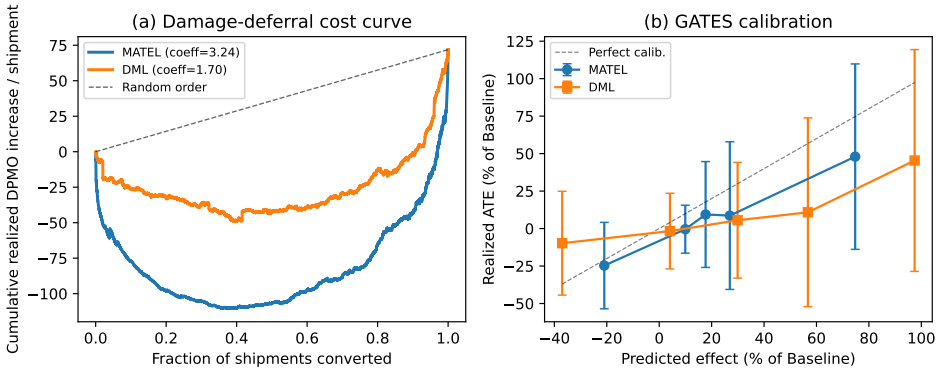


Figure D.1: PolyBag-B vs. Box-A. **(a)** Damage-deferral cost curve: MATEL's conversion ordering incurs less realized DPMO than DML and random at every volume. **(b)** GATES calibration (Box-A= 100): MATEL's group effects track the identity line; DML's are compressed. Bands are 95% CIs.

Table D.1: Pair-specific observational diagnostics versus Box-A. A BLP slope near 1 is favorable, while a CI excluding 0 indicates detectable heterogeneity. The deferral coefficient summarizes damage deferred relative to random ordering (higher is better). Relative ESS is normalized by the maximum ESS across reported contrasts. †: DML slope not significant ($p > 0.05$).

	BLP $\hat{\beta}$		Deferral coeff.		
Contrast (vs. Box-A)	MATEL	DML	MATEL	DML	Rel. ESS (%)
Product Packaging	1.01	0.66	**3.92**	2.71	41.2
Box-B	1.07	0.07†	**0.56**	0.17	14.0
Box-C	0.79	0.00†	**0.38**	0.10	5.0
Paper Mailer-A	1.09	1.02	**1.00**	0.50	47.0
Paper Mailer-B	0.96	0.01†	**0.56**	0.32	9.3
Paper Mailer-C	1.05	2.24	**0.73**	0.55	37.4
Paper Mailer-D	1.01	0.78	**0.80**	0.39	30.9
PolyBag-A	1.09	1.34	**0.70**	0.47	20.3
PolyBag-B	0.96	0.44	**3.24**	1.70	100.0

Table D.2: Simulated conversions and savings vs. the baseline scenario, as the damage-cost weight varies. Converted = Improved + Harmed; in this production-calibrated scenario, Box-A aggregates all corrugated formats; net reduction = outflow − inflow.

Damage-cost weight	Converted (% shipments)			Network-cost improvement (pp)	Box-A (%)		
	Improved	Harmed	Total		Outflow	Inflow	Net
0	17.63	1.31	18.94	2.71	24.86	0.11	24.75
0.5	20.27	1.11	21.39	3.22	24.86	0.11	24.75
1	25.16	0.97	26.13	3.74	24.85	0.13	24.72
2	24.87	10.40	35.27	3.15	24.83	0.20	24.63
5	24.03	21.39	45.42	2.96	24.62	0.64	23.98
10	23.01	25.11	48.12	1.94	23.90	2.06	21.84

Table D.3: Full population counterfactual DPMO predictions vs. observed rates, on a one-year rolling horizon. Difference between the predicted and observed indicates the severe selection bias in real-world data. In particular, boxes that provide high protections are observed with higher concession, whereas MATEL's estimations debias it. All values are *indexed to the Box-A observed rate* $= 100$ within each outcome (Predicted and Observed share the same normalization, so calibration is preserved). Residual is Predicted − Observed, in index points.

	Concession			**Problem Solve**		
Pack Type	Predicted	Observed	Residual	Predicted	Observed	Residual
Product Packaging	77.5	66.4	11.1	347.5	230.7	116.8
Box-A	61.3	100.0	−38.7	54.6	100.0	−45.4
Box-B	66.8	126.1	−59.3	238.3	267.0	−28.7
Box-C	63.2	91.1	−27.9	293.6	314.8	−21.2
Paper Mailer-A	73.1	80.7	−7.6	240.1	265.2	−25.1
Paper Mailer-B	63.9	88.8	−24.9	159.7	195.0	−35.3
Paper Mailer-C	87.1	63.7	23.4	410.4	356.5	53.9
Paper Mailer-D	69.4	82.8	−13.4	198.3	208.1	−9.8
PolyBag-A	77.6	76.6	1.0	160.3	139.6	20.7
PolyBag-B	80.2	70.8	9.4	285.4	168.4	117.0

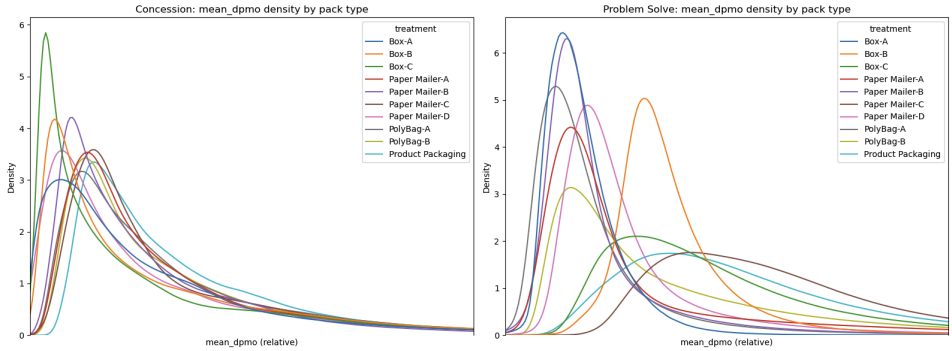


Figure D.3: Concession (left) and PS (right) counterfactual DPMO predictions by MATEL in a rolling horizon

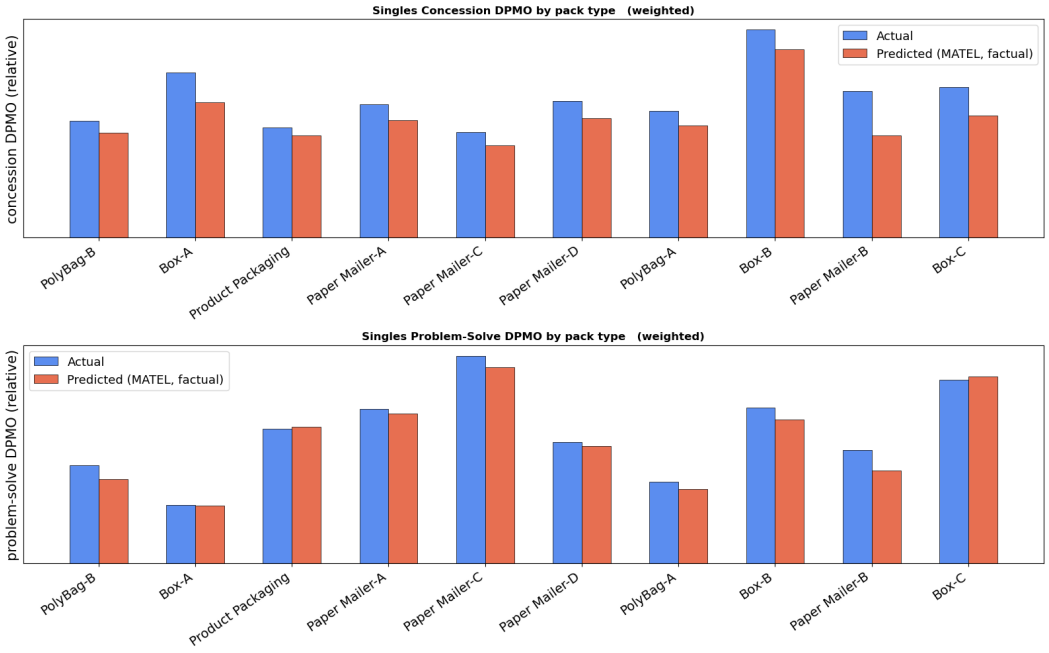


Figure D.2: Concession (top) and PS (bottom) DPMO distributions for 11M products (7B shipments) set by pack types, factual predictions by MATEL

Table D.4: Backbone comparison on held-out test sets: factual-prediction MAE improvement (%) over the baseline at three shipment-volume thresholds, and training time.

	MAE Improvement (%) ↑ (> shipments)			
Backbone	> 0	$> 1{,}000$	$> 10{,}000$	Train time
XGBoost	42.7	36.5	24.4	0.3h
ResNet	41.2	37.6	30.1	6h
FT-Transformer	44.3	39.3	28.9	13h

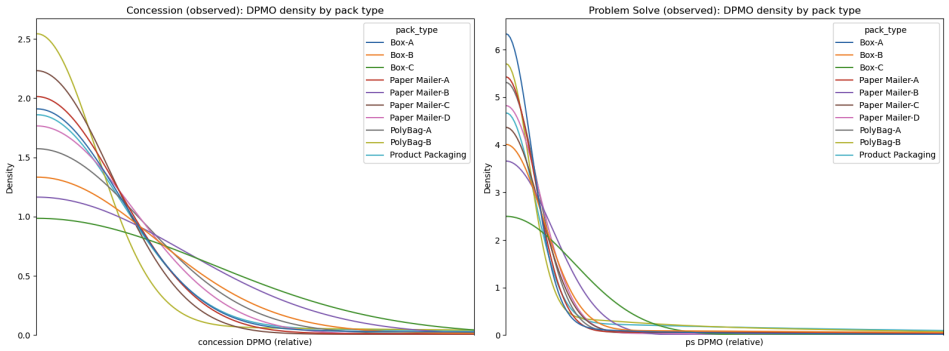


Figure D.4: Concession (left) and PS (right) observational DPMOs in a rolling horizon

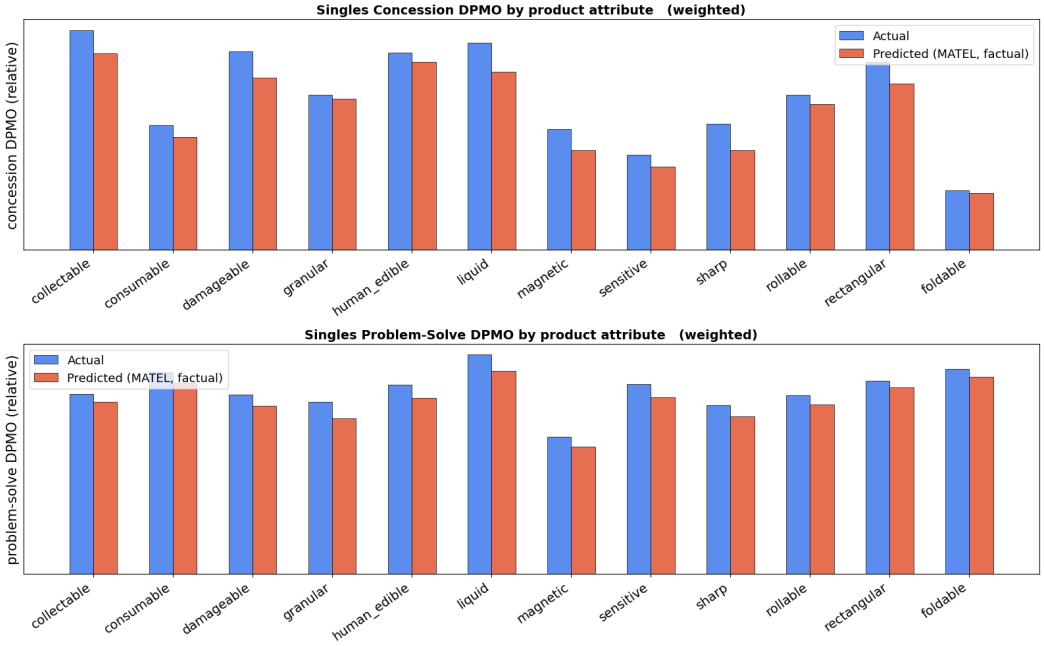


Figure D.5: Observed and MATEL-predicted concession (top) and PS (bottom) DPMO for 11M products and 7B shipments, grouped by the 12 product attributes defined in Table B.1.

Table D.5: Model hyperparameters for the backbone comparison (Table D.4), singles. All models use a Tweedie objective ($p = 1.5$); the deep backbones (Optuna-tuned) use PLE numerical embeddings and an AdamW optimizer. Depth, width, and batch size were included in the Optuna search ($n_{\text{blocks}} \in 2, 3, 4, 6, 10$, $d_{\text{block}} \in 128, 256, 512, 1024$, batch $\in 1024, 2048, 4096, 8192$); increasing depth or width, or shrinking the batch, lowered training loss but degraded held-out MAE by overfitting the heavy-tailed tail cells, so the tuned optimum favors the parsimonious $n_{\text{blocks}} = 3$, $d_{\text{block}} = 256$, batch $= 4096$.

Hyperparameter	Value
XGBoost	
n_estimators	2000
max_depth	8
learning_rate	0.05
early_stopping_rounds	500
tree_method	hist
objective	reg:tweedie
tweedie_variance_power	1.5
enable_categorical	true
ResNet-PLE	
n_blocks	3
d_block	256
PLE d_embedding	32
d_hidden_multiplier	2.0
dropout1 / dropout2	0.3 / 0.2
optimizer	AdamW
learning rate	3×10^{-4}
weight decay	5×10^{-4}
batch size	4096
FT-Transformer-PLE	
n_blocks	3
d_block	256
attention heads	8
FFN d_hidden multiplier	4/3
attention dropout	0.101
FFN dropout	0.173
residual dropout	0.022
optimizer	AdamW
learning rate	1.51×10^{-4}
weight decay	4.64×10^{-5}
batch size	2048

Table D.6: Pairwise ATE sign-direction accuracy for Singles concession on the 287k-product validation set. Each entry is the exposure-weighted average of the pairwise ATE over products with at least $n_{\min} = 10$ shipments in both arms of that pair. This pair-level diagnostic is not constrained to be transitive since the qualifying product set differs by pair. *Observed* is the observed sign of the pairwise ATE; *Est.* columns give each method's predicted sign; *Agree* marks agreement with *Observed*. MATEL recovers the observed-direction agreement on 41/45 pairs (91%) versus 28/45 (62%) for the DML benchmark.

		MATEL		DML	
Pair	**Observed**	**Est.**	**Agree**	**Est.**	**Agree**
PolyBag-A vs Paper Mailer-D	+	+	✓	−	✗
Paper Mailer-A vs Paper Mailer-D	−	−	✓	−	✓
Paper Mailer-A vs PolyBag-A	−	−	✓	−	✓
Paper Mailer-B vs Paper Mailer-D	−	−	✓	−	✓
Paper Mailer-B vs PolyBag-A	−	−	✓	+	✗
Paper Mailer-B vs Paper Mailer-A	−	−	✓	−	✓
Box-A vs Paper Mailer-D	−	−	✓	−	✓
Box-A vs PolyBag-A	−	−	✓	−	✓
Box-A vs Paper Mailer-A	−	−	✓	−	✓
Box-A vs Paper Mailer-B	−	−	✓	+	✗
Product Packaging vs Paper Mailer-D	−	−	✓	−	✓
Product Packaging vs PolyBag-A	−	−	✓	+	✗
Product Packaging vs Paper Mailer-A	+	+	✓	−	✗
Product Packaging vs Paper Mailer-B	−	+	✗	−	✓
Product Packaging vs Box-A	+	+	✓	+	✓
Box-B vs Paper Mailer-D	−	−	✓	+	✗
Box-B vs PolyBag-A	−	−	✓	−	✓
Box-B vs Paper Mailer-A	−	−	✓	−	✓
Box-B vs Paper Mailer-B	−	−	✓	+	✗
Box-B vs Box-A	+	+	✓	+	✓
Box-B vs Product Packaging	+	−	✗	+	✓
Paper Mailer-C vs Paper Mailer-D	+	+	✓	+	✓
Paper Mailer-C vs PolyBag-A	−	+	✗	−	✓
Paper Mailer-C vs Paper Mailer-A	+	+	✓	+	✓
Paper Mailer-C vs Paper Mailer-B	+	+	✓	−	✗
Paper Mailer-C vs Box-A	+	+	✓	+	✓
Paper Mailer-C vs Product Packaging	+	+	✓	−	✗
Paper Mailer-C vs Box-B	+	+	✓	+	✓
PolyBag-B vs Paper Mailer-D	+	+	✓	−	✗
PolyBag-B vs PolyBag-A	−	−	✓	+	✗
PolyBag-B vs Paper Mailer-A	+	+	✓	+	✓
PolyBag-B vs Paper Mailer-B	+	+	✓	−	✗
PolyBag-B vs Box-A	+	+	✓	+	✓
PolyBag-B vs Product Packaging	−	−	✓	−	✓
PolyBag-B vs Box-B	+	+	✓	+	✓
PolyBag-B vs Paper Mailer-C	−	−	✓	−	✓
Box-C vs Paper Mailer-D	−	−	✓	+	✗
Box-C vs PolyBag-A	−	−	✓	+	✗
Box-C vs Paper Mailer-A	−	−	✓	−	✓
Box-C vs Paper Mailer-B	−	−	✓	−	✓
Box-C vs Box-A	+	+	✓	−	✗
Box-C vs Product Packaging	+	−	✗	+	✓
Box-C vs Box-B	+	+	✓	+	✓
Box-C vs Paper Mailer-C	−	−	✓	+	✗
Box-C vs PolyBag-B	−	−	✓	+	✗
Total correct		**41/45 (91%)**		**28/45 (62%)**	

Table D.7: **Estimator configurations in the synthetic study.** Hyperparameters for every estimator were tuned by Optuna for the neural architectures and by grid search for the tree learners. Tree learners are XGBoost (learning rate 0.05), written as trees$\times$depth. All estimators see the same feature set; for DML, whose linear final stage fits one coefficient per column per arm, the confounder block enters as effect modifiers and the remaining features as controls. Both neural estimators combine five initializations. MATEL's Tweedie heads output $\log \mu$, so an unconverged initialization can emit a few per-row contrasts far outside the outcome range, and initializations are combined by median. MATEL's propensity clip $e \in [0.01, 0.99]$ never binds under Assumption 4; DML floors its treatment residual at 10^{-5}, which leaves the inverse weights effectively unbounded. Fitting the same honest forest on cross-fitted residuals (econml `CausalForestDML`) rather than on raw outcomes raises its $\sqrt{\text{PEHE}}$ by 200–400 DPMO at every overlap level, the cost Lemma 2 attributes to residualization.

Neural estimators

	MATEL	**DragonNet (vanilla)**
Backbone	ResNet-PLE, 3 blocks $\times$ 192, target-supervised bins	MLP, 2 layers $\times$ 128
Outcome heads	Tweedie, p=1.5	MSE, 64 hidden
Propensity	hybrid encoder on W	shared representation
Targeting	pairwise ε, reconciled	per-pair εH
Optimizer	AdamW 10^{-3}, one-cycle, $\leq$ 100 epochs, patience 25	Adam 10^{-3}, 300 epochs
Clips	C_ε=100, $e \in [0.01, 0.99]$, λ=10^{-2}, $\lvert\tau^*\rvert \leq y_{\text{clip}}$	—

Tree-based estimators

	Causal Forest	**X-learner**
Outcome model	—	Tweedie, 300$\times$6
Propensity	—	logistic, 300
Final stage	honest forest, 300 trees, leaf $\geq$ 100	pseudo-Huber, 300$\times$6

Double machine learning

	DML (linear)
Outcome model	Tweedie, 500$\times$8
Propensity	logistic, 500
Final stage	linear, 5-fold
Residual floor	$\lvert T - \hat{g}\rvert \geq 10^{-5}$ (library default)

E SYNTHETIC DATA-GENERATING PROCESS

The DGP provides known ground truth for every arm, so estimation error is measured directly. Because the benchmarks report pairwise contrasts, all methods are compared on that quantity: each arm against a common baseline. Transitivity and coherent per-arm levels are not scored here. MATEL obtains them while targeting each pair independently, which is what keeps the correction weights bounded under thin relative overlap. Proposition 4 establishes them analytically.

Covariates are real product rows in our runs, but any table supplying the following fields below can be used to substitute the original product rows; treatment and outcome are simulated using the procedure described in this section. Write W for the product-level block (dimensions, price, rating, attributes), X for the operational environment, and $Z = (W, X)$. Let $\tilde{w}, \tilde{v}, \tilde{p}$ be weight, volume and price standardized on the training split and clipped to $[-3, 3]$, and $d \in \{0, 1\}$ the damageable flag. All constants appear in Table E.1.

Data sample. Whole products are sampled to the target size and partitioned in $64/16/20$ at product level. Standardization moments and the propensity model are estimated on the training split only.

Outcome surfaces. A single baseline surface is specified in closed form and arm effects are imposed on top of it,

$$\mu_0(z) = 10^3 \exp\big\{\beta_0 + \beta_1\tilde{w} + \beta_2\tilde{v} + \beta_3\tilde{p} + \beta_4\tilde{w}\tilde{v} + \beta_5 d\big\}, \qquad \mu_t(z) = \mu_0(z)\exp\big\{a_t + b_t\, s(z)\big\}, \tag{62}$$

with μ_0 in DPMO and $s(z)$ the standardized average of $(\tilde{w}, \tilde{v}, \tilde{p})$. The base arm has $a_b = b_b = 0$; the other four take a_t equally spaced on $[-\kappa_a, \kappa_a]$ and b_t on $[\kappa_h, -\kappa_h]$, fixed across seeds so every replicate targets the same estimand. Effects are therefore multiplicative in the baseline risk, and their sign can vary across products through $s(z)$.

Treatment Assignment. For the focal pair (k, b) the assignment is written as a pair mass times a within-pair split,

$$g_k = q\, e_{k,b}, \qquad g_b = q\,(1 - e_{k,b}), \qquad e_{k,b}(w) = \frac{1}{1 + \rho^{u(w)}}, \quad u(w) = \tfrac{1}{2}\{1 + \tanh\tilde{v}\}, \tag{63}$$

with the remaining mass $1 - q$ spread over the other arms in proportion to a multinomial logit fitted on W. The pair odds $g_b/g_k = \rho^u$ lie in $[1, \rho]$ for every w, so Assumption 4 holds with the stated ρ at every level of the sweep. The lightest c of products receive pair mass $q_{\min}$; the rest receive $q_0 = (\bar{q} - c\, q_{\min})/(1 - c)$, chosen so the average pair mass is $\bar{q}$ whatever $q_{\min}$ is. The focal pair therefore contributes the same expected number of rows at every level, and only its concentration across products changes. Lowering $q_{\min}$ makes the pair rare for a minority of products without shrinking the sample or disturbing the balance between the two arms. An arm's own probability is the pair mass times its share within the pair, and that share never falls below $1/(1 + \rho)$, so the absolute margin (18) is $\eta = q_{\min}/(1 + \rho) \approx 0.30\, q_{\min}$: the three levels place η at 3.0×10^{-2}, 3.0×10^{-3}, and 3.0×10^{-4}.

Outcomes. Given $T_i \sim \text{Categorical}\{g(w_i)\}$,

$$Y_i \mid Z_i, T_i{=}t \;\sim\; \text{Tweedie}_p\big(\mu_t(z_i), \phi\big), \qquad p = 1.5, \tag{64}$$

sampled as a Poisson number of Gamma jumps, so $\mathbb{P}(Y{=}0) = \exp(-\lambda)$ with $\lambda = \mu^{2-p}/\{\phi(2{-}p)\}$. The dispersion ϕ is chosen so that the average zero probability matches the observed zero fraction of concession; averaging over units rather than evaluating at $\overline{\mu}$ matters because μ is right-skewed. A Gamma or log-normal response cannot reproduce the point mass at zero that dominates the observed column.

Targets. Since (62) is frozen, $\tau_{k,b}(z) = \mu_k(z) - \mu_b(z)$ and $\overline{\tau}_{k,b} = \mathbb{E}_P[\tau_{k,b}(Z)]$ are exact values. Methods are retrained from scratch on (Z, T, Y) without receiving $\{\mu_t\}$. Errors are computed on the held-out split over all contrasts; ranking metrics are computed on the pairwise overlap subset $e_{k,b} \in [0.1, 0.9]$, since incremental gain from targeting is defined only where a product could receive either arm.

Algorithm 1 Synthetic data generation

Require: arm set with base b, target size n, pair-mass floor $q_{\min}$, seed
1: sample whole products to n rows; split $64/16/20$ by product
2: standardize $(\tilde{w}, \tilde{v}, \tilde{p})$ using training-split moments; form $s(z)$
3: build μ_0 and $\{\mu_t\}$ from (62) ▷ identical across seeds
4: fit the multinomial logit on W over training rows
5: set $e_{k,b}$ from (63); set $q = q_{\min}$ on the lightest c of products, q_0 elsewhere
6: assemble g from (63) and renormalize
7: draw $T_i \sim \text{Categorical}(g_i)$
8: solve for ϕ so the average zero probability matches the observed zero fraction
9: draw Y_i from (64)
10: **return** synthetic (T, Y), the frozen $\{\mu_t\}$, and the overlap subsets $\{e_{k,b} \in [0.1, 0.9]\}$

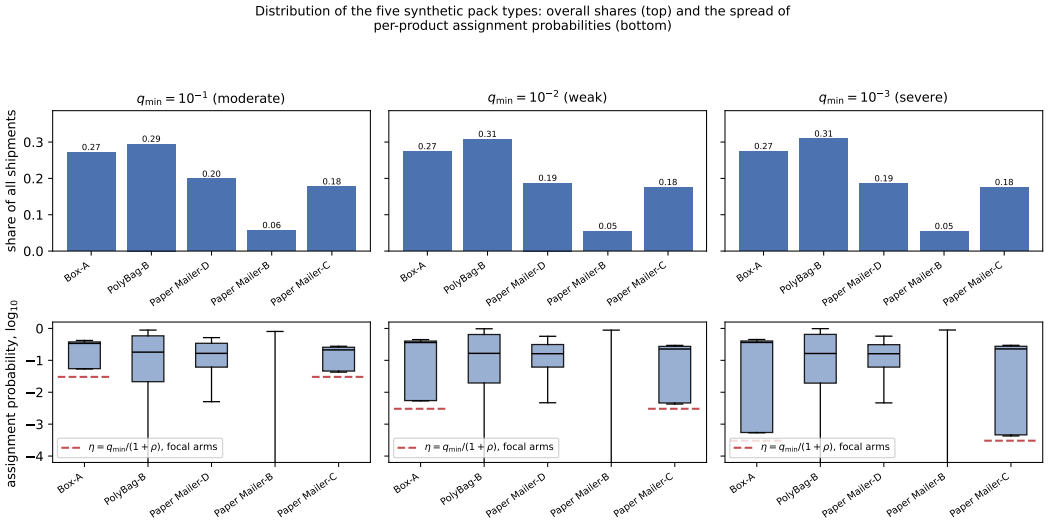


Figure E.1: **Distribution of the five synthetic pack types.** Top: share of synthetic units assigned to each pack type at each $q_{\min}$. Bottom: spread of per-product assignment probabilities; dashed segments mark the implied arm-level propensity floor $\eta = q_{\min}/(1+\rho)$ for the designated focal pair (Box-A and Paper Mailer-C), which defines the synthetic overlap regimes. Boxes span the interquartile range and whiskers the 1st to 99th percentiles. Arm names follow Section 2.

Table E.1: Constants of the generating process (left) and the quantities the sweep moves or holds fixed (right). κ_a places effects at roughly a third of the outcome level and ρ matches the fitted pairwise ratios, both taken from Table 1.

constant			$q_{\min}$=10^{-1}	10^{-2}	10^{-3}
$(\beta_0, \ldots, \beta_5)$	$(.80, .35, .25, -.15, .10, .30)$	$\mathbb{E}[1/g_k]$	9.7	69.3	667.9
$\kappa_a,\ \kappa_h$	$0.30,\ 0.30$	ρ	2.30	2.30	2.30
ρ	2.30	$\mathbb{E}[q]$	0.45	0.45	0.45
$\bar{q},\ c$	$0.45,\ 0.30$	$n_{k,b}$	26,923	26,933	26,985
p, zero fraction	$1.5,\ 0.90$	$\overline{\tau}_{k,b}$	504	504	504