Causal Multi-Task Demand Learning

Varun Gupta
Dept. of Operations and Information Systems
University of Utah
varun.gupta@eccles.utah.edu

Vijay Kamble
Dept. of Information and Decision Sciences
University of Illinois Chicago
kamble@uic.edu

Abstract

We study multi-task demand learning for retail pricing, where firms seek to estimate heterogeneous linear price-response functions across many decision contexts. Each context has rich covariates but limited price variation, motivating transfer across tasks. However, such transfer is complicated by endogeneity: prices may be arbitrarily correlated with unobserved task-level demand determinants.

We propose a meta-learning framework that identifies the conditional mean of task-specific causal demand parameters despite this confounding, provided each task contains at least two distinct, locally exogenous price points. The key idea is to carefully design the information available to the learner. We condition on the full price history to address cross-task confounding, but withhold the demand outcomes associated with two locally exogenous price points to preserve identifiability. These withheld outcomes provide supervision through a randomized query.

We establish that this information design is maximally uniformly valid for the squared-loss objective: any nonconstant refinement of the conditioning set that reveals information about the withheld outcomes or the query index admits a data-generating process under which the objective no longer identifies the conditional mean causal target. Synthetic experiments demonstrate improved estimation of causal demand parameters under strong confounding, while retail experiments show improved held-out demand prediction relative to standard transfer-learning baselines.

1 Introduction

Large retail firms routinely set prices across a wide range of selling contexts. These contexts may correspond to a single product offered across geographically dispersed stores and channels within an omnichannel network (context is the geography and store-level features) or to distinct products within an e-commerce catalog (context is the product features). In all cases, each decision environment is shaped by heterogeneous demand drivers, including customer demographics, competitive conditions, local preferences, and product-specific attributes. Consequently, effective pricing requires accurate estimation of context-specific price-response functions.

Estimating such heterogeneous demand responses presents a statistical challenge. Within any given store or product, price variation is typically sparse: operational frictions often limit exposure to only a small set of prices over extended periods. In contrast, the cross-sectional dimension is large, with many stores or products and rich covariate information. This imbalance naturally motivates multi-task learning or partial pooling approaches [Caruana, 1997, Baxter, 2000, Gelman et al., 2013], in which a shared model maps observable characteristics to task-specific demand parameters, borrowing strength across tasks to compensate for limited within-task variation.

However, many factors that determine prices are unobserved or only partially observed by the econometrician. As a result, prices may remain systematically correlated with the unobserved demand determinants across tasks even after conditioning on the observed covariates, leading to

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

unobserved confounding. This raises a central question: can cross-task transfer learning improve estimation of causal demand parameters despite such confounding, and under what conditions?

We formalize a model of multi-task pricing with heterogeneous linear price responses, where observed prices may depend on latent demand determinants. Standard pooled and meta-learning estimators generally converge to biased policy-dependent estimands rather than causal targets in this model. We make the following main contributions:

1. We propose a new information design principle for meta-learning in this setting, termed Decision-Conditioned Masked-Outcome Meta-Learning (DCMOML). Assuming at least two distinct locally exogenous price points per task, we show that DCMOML identifies and consistently estimates the conditional mean of task-specific causal demand parameters given a specified information set. This information set includes all task prices to address confounding, while masking two demand observations corresponding to locally exogenous prices to address the identifiability issues that arise when prices are fully observed by the meta-learner. Crucially, supervision is provided by query randomization over the two obfuscated demand outcomes.
2. We show that the DCMOML information set is maximally uniformly valid for its two-point squared-loss supervision: for every nonconstant refinement that reveals withheld-outcome or query-index information, some admissible data-generating process makes the refined population loss fail to recover the conditional mean target uniquely.
3. On synthetic data with known parameters, DCMOML approaches the Bayes floor of its information set and outperforms pooled regression, empirical Bayes, and standard support–query meta-learning under strong confounding; standard meta-learning can be more accurate for intercepts when confounding is weak. On real retail data, DCMOML improves prediction at withheld prices.

We note that task-level identification is not the primary difficulty in our setting. Under our model, two distinct locally exogenous price points within a task already supply two valid conditional moment equations for the causal parameter, and hence a two-point OLS estimate of it. However, such estimates are highly variable when each task contains only a small number of observations, and may lack finite moments when the two prices are close. Our goal is to improve precision by borrowing strength across tasks through transfer learning. And this is precisely where confounding becomes consequential.

If each task contained only a single price point, the problem would collapse to a standard regression with endogenous prices, where unbiased estimation is generally impossible without additional assumptions or valid instruments. Our main insight is that locally exogenous price variation within each task allows us to use cross-task learning to improve precision, without requiring prices to be exogenous across tasks.

Overall, our contribution bridges empirical demand estimation under endogenous pricing with cross-task learning in data-sparse environments. Improved estimation of these context-specific causal demand primitives can enable more principled downstream pricing decisions in retail operations, hence our contribution is of practical value.

2 Related Work

Demand estimation under endogenous prices. A central concern in empirical demand estimation is that prices are often chosen in response to demand information unobserved by the econometrician. Classical approaches address this endogeneity using instruments, control functions, panel variation, or structural equilibrium restrictions [Berry, 1994, Berry et al., 1995, Hausman, 1996, Nevo, 2001, Petrin and Train, 2010, Berry and Haile, 2014, Angrist and Pischke, 2009, Wooldridge, 2010]. DeepIV extends the instrumental-variables approach to flexible nonlinear prediction by using instruments to generate conditionally exogenous treatment variation [Hartford et al., 2017]. We assume neither excluded instruments nor conditional independence of prices from latent demand determinants. Instead, DCMOML leverages repeated, locally exogenous decisions within each task to identify a conditional mean causal target in a many-task, few-observation regime.

Partial pooling, empirical Bayes, and meta-learning. Hierarchical Bayes, random-effects models, and empirical Bayes shrinkage reduce the variance of task-level estimation when each unit has few observations [Robbins, 1956, Laird and Ware, 1982, Gelman et al., 2013, Efron, 2010]. These approaches typically rely on valid causal estimates at the task level, which our model provides through

the two locally exogenous prices per task. In standard formulations with a price-independent task-parameter prior, EB/random-effects methods use realized prices only as regressors in the outcome likelihood. DCMOML additionally treats the price history as information about latent demand parameters and learns this relationship flexibly, without specifying a model of how prices are set. Multi-task and meta-learning methods are natural for demand learning with many stores or products and sparse within-task price variation [Caruana, 1997, Baxter, 2000, Evgeniou and Pontil, 2004, Finn et al., 2017, Hospedales et al., 2022]. However, when prices encode latent demand information, a pooled predictor or standard support–query meta-learner may fail to learn the causal price response. DCMOML is designed to precisely address this issue.

Causal machine learning. Causal trees and forests, double/debiased machine learning (DML), and R-learners estimate treatment effects using flexible regression and, in some formulations, orthogonalization [Athey and Imbens, 2016, Wager and Athey, 2018, Chernozhukov et al., 2018, Nie and Wager, 2021, Imbens and Rubin, 2015]. Their observed-confounder formulations rely on assumptions such as unconfoundedness and overlap. DML also accommodates unobserved confounding when valid instruments are available. Our setting assumes neither that observed covariates capture all confounding nor that excluded instruments are available; identification instead relies on the linear demand structure and repeated, locally exogenous observations within tasks.

Correlated random effects and random-coefficient panels. Our setting is closely related to correlated random-effects and random-coefficient panel models, in which unit-level heterogeneity may be correlated with the within-unit regressor history [Swamy, 1970, Mundlak, 1978, Chamberlain, 1982, 1984, Wooldridge, 2010]. Mundlak–Chamberlain approaches relax the standard random-effects independence assumption by modeling the conditional mean of latent unit effects given the regressor history, often through linear covariate summaries. In these models, outcomes enter the estimating equations, while the conditioning object is typically fixed in advance as a function of the observed regressors. DCMOML adopts this conditional-random-coefficient perspective but differs in what conditioning information is targeted and how the target is learned. In our meta-learning setting, information about the latent demand type comes from two sources: the endogenous price history, which reflects prior pricing decisions, and the within-task relationship between realized prices and demands, which is informative about the task's random coefficient. This necessitates that (a) the demand outcomes enter the conditioning information, and (b) the targeted conditional mean be a flexible function of prices and outcomes as opposed to linear price summaries. These aspects create a unique identification challenge that we address using outcome masking and query randomization.

3 Model

Tasks, Data, and Notation. We study a multi-task demand learning setting with N tasks indexed by $i = 1, \ldots, N$. Each task can be interpreted as a distinct store selling a fixed product. Tasks are heterogeneous and described by observable covariates $Z_i \in \mathbb{R}^d$, capturing demographics, competition, geography, and other store-level features. We assume $\{Z_i\}_{i=1}^N$ are i.i.d. draws from an unknown distribution $\mathcal{P}_Z$. For each task i, we observe a sequence of K price–demand pairs

$$\{(p_{ik}, D_{ik})\}_{k=1}^K,$$

where $p_{ik} \in \mathbb{R}$ is the posted price and $D_{ik} \in \mathbb{R}$ is realized demand. We focus on regimes with limited within-task variation, where K is small (e.g., $K = 2$ or 3).

Structural Demand Model. Demand is assumed linear in price within each task:

$$D_{ik} = \theta_i^0 + \theta_i^1 p_{ik} + \epsilon_{ik}, \tag{1}$$

where $\Theta_i = (\theta_i^0, \theta_i^1)^\top \in \mathbb{R}^2$ are task-specific demand parameters and ϵ_{ik} is an idiosyncratic shock. Let $P_{ik} = (1,\, p_{ik})^\top$, so that

$$D_{ik} = P_{ik}^\top \Theta_i + \epsilon_{ik}.$$

We assume the shocks have mean zero and finite variance, and that $\epsilon_{i1}, \ldots, \epsilon_{iK}$ and (Z_i, Θ_i) are mutually independent.

Heterogeneity and Shared Structure. We decompose task-specific parameters as

$$\Theta_i = g(Z_i) + \Omega_i, \tag{2}$$

where $g : \mathbb{R}^d \to \mathbb{R}^2$ is an unknown function capturing systematic variation across tasks, and $\Omega_i \in \mathbb{R}^2$ is an idiosyncratic task-specific deviation. We assume $\mathbb{E}[\Omega_i \mid Z_i] = 0$ and that $\{\Omega_i\}$ are i.i.d. with finite second moments. Substituting into (1) yields

$$D_{ik} = P_{ik}^\top g(Z_i) + P_{ik}^\top \Omega_i + \epsilon_{ik}. \tag{3}$$

Task tuples $(Z_i, \Omega_i, p_{i1:K}, \epsilon_{i1:K})$ are i.i.d. across tasks.

Price Assignment and Confounding. We allow the task-specific prices $\{p_{ik}\}_{k=1}^K$ to depend on the unobserved heterogeneity Ω_i, even after conditioning on the observed context Z_i. Thus, prices may be endogenous through their dependence on Ω_i across tasks. We refer to this feature as *cross-task confounding*.

At the same time, we assume that there are at least two distinct *locally exogenous* prices in each task, which supply two valid conditional moment equations for Θ_i at the task level:

Assumption 1 (Local exogeneity)**.** *For each task i, there exist two indices a_i and b_i, measurable with respect to $(Z_i, p_{i1:K})$, such that $p_{ia_i} \neq p_{ib_i}$ almost surely. Moreover, for each $k \in \{a_i, b_i\}$,*

$$\mathbb{E}\big[\epsilon_{ik} \mid Z_i, p_{i1:K}, \mathbf{D}_i^{-k}\big] = 0, \tag{4}$$

where $\mathbf{D}_i^{-k} = \{D_{ij} : j \neq k\}$ denotes the task history with the kth demand observation omitted.

For example, Assumption 1 holds when prices are pre-committed, i.e., $p_{i1:K}$ is set before the demand shocks are realized, possibly depending arbitrarily on (Z_i, Ω_i) and on randomization independent of the shocks. Prices are then endogenous through latent demand type, yet each mean-zero shock ϵ_{ik} is independent of $(Z_i, \Omega_i, p_{i1:K}, \epsilon_{i,-k})$, and hence of $(Z_i, p_{i1:K}, \mathbf{D}_i^{-k})$. Hence the assumption holds at every index, and any two indices with distinct prices can serve as (a_i, b_i).

Local exogeneity can also hold after an adaptive history: if each task ends at a fixed pair of distinct probe prices chosen from past demand before either probe outcome is observed, as in Kiefer–Wolfowitz experimentation [Kiefer and Wolfowitz, 1952], that final pair is locally exogenous under the model's shock-independence assumptions.

Assumption 2 (Moment condition)**.** *The task parameters and the noiseless demands at the observed prices have finite second moments:*

$$\mathbb{E}\|\Theta_i\|_2^2 < \infty, \qquad \mathbb{E}\big[(P_{ik}^\top \Theta_i)^2\big] < \infty \quad \textit{for } k = 1, \ldots, K. \tag{5}$$

These conditions are mild; for example, they hold whenever the task parameters and prices are almost surely bounded.

Causal Target and Objective. The causal object of interest is the task-specific parameter Θ_i, which governs demand responses under counterfactual price interventions. Under Assumption 1, the two locally exogenous observations supply valid conditional moment equations for Θ_i, and determine a two-point OLS estimate of it. With a small K, however, that estimate is noisy and unstable, and its unconditional moments need not exist when the two prices are close. To obtain a less noisy estimate, in the spirit of meta-learning, our objective is to leverage the cross-task distributional information available in the many-tasks regime to target the conditional expectation of the causal parameter Θ_i given task-level information sets of observables, where the key challenge is confounding.

4 Baseline Transfer-Learning Approaches and Challenges

We first discuss how standard pooling and meta-learning approaches are biased in our setting.

Shared-Model Learning. The simplest transfer learning approach is to ignore task-level heterogeneity and estimate only the shared mapping $g(\cdot)$ by pooling all observations. A natural baseline in this spirit is to posit a parameterized function class $\{g_\Lambda(\cdot)\}_{\Lambda \in \mathcal{H}}$ (e.g., deep neural networks), and to estimate Λ by pooling all observations across tasks:

$$\widehat{\Lambda} \in \arg\min_{\Lambda \in \mathcal{H}} \sum_{i=1}^N \sum_{k=1}^K \big(D_{ik} - P_{ik}^\top g_\Lambda(Z_i)\big)^2. \tag{6}$$

This is appealing in the many-tasks regime, as it avoids noisy within-task estimation, but it is generally inconsistent under our model. Substituting the data-generating process gives

$$D_{ik} - P_{ik}^\top g(Z_i) = P_{ik}^\top \Omega_i + \epsilon_{ik}.$$

Because prices are endogenously assigned, P_{ik} may be correlated with Ω_i, implying $\mathbb{E}[P_{ik}^\top \Omega_i \mid Z_i] \neq 0$, violating the orthogonality condition required for consistency of (6): a classic endogeneity failure. Appendix A gives an example in which confounded near-optimal pricing reverses the sign of the pooled slope.

Meta-Learning. Standard support–query meta-learning attempts to leverage the benefits of pooling while allowing for per-task adaptation [Hospedales et al., 2022, Finn et al., 2017]. Let $g_\Lambda : (Z_i, \mathcal{S}_i) \mapsto \widehat{\Theta}_i$, where $\mathcal{S}_i = \{(P_{i1}, D_{i1}), \ldots, (P_{i,K-1}, D_{i,K-1})\}$ is a *support set*. The final observation (P_{iK}, D_{iK}) is treated as a *query*, and Λ is estimated via

$$\widehat{\Lambda} \in \arg\min_{\Lambda} \sum_{i=1}^{N} \left(D_{iK} - P_{iK}^\top g_\Lambda(Z_i, \mathcal{S}_i)\right)^2. \tag{7}$$

While one might expect this to recover $m_i \triangleq \mathbb{E}[\Theta_i \mid Z_i, \mathcal{S}_i]$, this is not the case in general, since endogeneity leads to inconsistent estimation. For instance, if the query index is locally exogenous, the expected residual of m_i is

$$\mathbb{E}\big[D_{iK} - P_{iK}^\top m_i \;\big|\; Z_i, \mathcal{S}_i\big] = \operatorname{tr} \operatorname{Cov}(P_{iK}, \Omega_i \mid Z_i, \mathcal{S}_i),$$

which need not be zero, because p_{iK} can remain correlated with Ω_i even after conditioning on $(Z_i, \mathcal{S}_i)$. And whenever it is nonzero, m_i is not a minimizer of the objective in (7). Thus, meta-learning generally converges to a policy-dependent estimand. In Appendix A, we present an example of this failure in the same setting where shared-model learning fails.

5 Causal Identification

We formulate causal identification of a conditional target as a meta-learning problem in which both the support information and query supervision must be carefully designed to address confounding. We develop this construction step by step before presenting our estimator.

Conditioning on the Full Price History. Suppose the query index K is locally exogenous. Because prices are confounded with latent task parameters, we first enlarge the standard support–query conditioning set of meta-learning to include the full price history. Specifically, consider the target

$$\mathbb{E}[\Theta_i \mid Z_i,\ p_{i1}, \ldots, p_{iK},\ D_{i1}, \ldots, D_{i,K-1}], \tag{8}$$

which conditions on all endogenous price realizations, including the query price p_{iK}, leaving the demand D_{iK} as supervision target for learning. A natural meta-learning estimator solves

$$\widehat{\Lambda} \in \arg\min_{\Lambda} \sum_{i=1}^{N} \left(D_{iK} - P_{iK}^\top g_\Lambda\big(Z_i,\ p_{i1:K},\ D_{i,1:K-1}\big)\right)^2, \tag{9}$$

with $P_{iK} = (1, p_{iK})^\top$. Because the full confounded price vector $p_{i1:K}$ is included as a covariate, this removes confounding at the level of conditional means, rendering the residual in (9) mean-zero under the target (8). However, it introduces a new failure mode of *identifiability*. The issue arises because the meta-learner has explicit access to the query price point p_{iK}, and the supervision signal for task i enters only through the scalar inner product $P_{iK}^\top \Theta_i$. Consequently, the empirical risk in (9) is invariant to shifts of the predicted parameter vector along directions orthogonal to P_{iK}. Formally, for any candidate prediction $\widehat{\Theta}_i$ and any measurable scalar function ϕ of the inputs, define

$$\widetilde{\Theta}_i \;\triangleq\; \widehat{\Theta}_i + \phi(\text{inputs}_i)[\,p_{iK}, -1\,]^\top.$$

Since $P_{iK}^\top [\,p_{iK}, -1\,]^\top = 0$, both $\widehat{\Theta}_i$ and $\widetilde{\Theta}_i$ achieve exactly the same objective value in (9). When g_Λ is expressive, many such solutions can exist, so this objective cannot uniquely select the target in (8).

Identification via Partial Obfuscation of the Decision History. We now construct an information set and supervision policy for a meta-learner that continues to condition on all prices to address

confounding, while restoring identification by preventing the learner from identifying the query regressor.

Step 1: Randomizing the query index. In standard meta-learning, the query index (e.g., $k = K$) is fixed, so the learner knows which price enters the loss. To avoid this, we randomize the query within a locally exogenous two-point subset. Specifically, select a locally exogenous index pair $A_i = \{a_i, b_i\} \subseteq [K]$ (which exists by Assumption 1), potentially as a function of $(Z_i, p_{i1:K})$. Draw $\kappa_i \sim \mathrm{Unif}\{a_i, b_i\}$ independently across tasks, and treat $(p_{i\kappa_i}, D_{i\kappa_i})$ as the query.

Step 2: Outcome masking. Randomization alone is insufficient: the learner can still infer κ_i because the observed demand vector is aligned with prices, leaving exactly one missing outcome. To prevent this, we mask both candidate query outcomes and instead target

$$\mathbb{E}\Big[\Theta_i \mid Z_i,\ p_{i1:K},\ \mathbf{D}_i^{-A_i}\Big], \tag{10}$$

where $\mathbf{D}_i^{-A_i} = \{D_{ij} : j \notin A_i\}$. Under this design, both p_{ia_i} and p_{ib_i} are unmatched with outcomes, so the query price $p_{i\kappa_i}$ is not measurable from the inputs. This restores the variation in the query price conditioned on the information set required for target identification. We next present our main estimator.

5.1 The estimator: Decision-Conditioned Masked-Outcome Meta-Learning

Definition 1 (Decision-Conditioned Masked-Outcome Meta-Learning (DCMOML))**.** Fix a hypothesis class $\{g_\Lambda : \Lambda \in \mathcal{H}\}$ with $g_\Lambda : \mathcal{Z} \times \mathbb{R}^K \times \mathbb{R}^{K-2} \to \mathbb{R}^2$, where the inputs correspond to $(Z_i, p_{i1:K}, \mathbf{D}_i^{-A_i})$. For each task i, draw κ_i according to Step 1, and form the masked-outcome information set

$$X_i^{-A_i} = \big(Z_i,\ p_{i1:K},\ \mathbf{D}_i^{-A_i}\big). \tag{11}$$

The *DCMOML estimator* is any empirical risk minimizer

$$\widehat{\Lambda}_{\mathrm{DCMOML}} \in \arg\min_{\Lambda \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{N} \Big(D_{i\kappa_i} - P_{i\kappa_i}^\top g_\Lambda\big(X_i^{-A_i}\big)\Big)^2, \tag{12}$$

where $P_{ik} = (1, p_{ik})^\top$. We refer to $g_{\widehat{\Lambda}_{\mathrm{DCMOML}}}$ as the *DCMOML meta-learner*.

Remark 1. Instead of sampling $\kappa_i \sim \mathrm{Unif}(A_i)$, one can minimize the *average* loss over the two candidate query indices. Both have the same population risk; averaging removes the randomness of the query draw, though the two empirical objectives need not share finite-sample minimizers. We use the averaged form in all experiments. Appendix E gives a design-weighted regression interpretation of this loss.

5.2 Main results: Identification, Consistency, and Maximal Validity

We now formalize the guarantees of the proposed information design. The first result establishes the identification of the causal target and gives consistency of the empirical procedure, and the second shows that the design is maximally uniformly valid. Proofs are provided in Appendix B.

Theorem 1 (Identification and consistency of DCMOML)**.** *Consider the model in Section 3. Suppose that Assumptions 1 and 2 hold, and let $A_i = \{a_i, b_i\}$ be any locally exogenous pair of indices chosen as a measurable function of $(Z_i, p_{i1:K})$. Let $\kappa_i \sim \mathrm{Unif}(A_i)$ independently across tasks and independently of all structural variables conditional on A_i, and let $X_i^{-A_i}$ denote the masked-outcome information set defined in* (11)*. For any measurable g satisfying $\mathbb{E}[(P_{i\kappa_i}^\top g(X_i^{-A_i}))^2] < \infty$, define the population risk*

$$\mathcal{L}(g) \triangleq \mathbb{E}\left[\Big(D_{i\kappa_i} - P_{i\kappa_i}^\top g(X_i^{-A_i})\Big)^2\right],$$

where the expectation is over the joint law induced by the pricing policy and the two-point design. Define $g^*(x) \triangleq \mathbb{E}[\Theta_i \mid X_i^{-A_i} = x]$. Then the following hold.

1. ***Identification.*** *The function g^* is the unique minimizer of $\mathcal{L}$: if $\mathcal{L}(g) = \mathcal{L}(g^*)$, then*

$$g(X_i^{-A_i}) = g^*(X_i^{-A_i}) \qquad \textit{almost surely}.$$

Consequently, $\mathbb{E}[\Theta_i \mid X_i^{-A_i}]$ is identified.

2. ***Consistency under realizability.*** *Let $\{g_\Lambda : \Lambda \in \mathcal{H}\}$ be the DCMOML hypothesis class and define $\mathcal{L}(\Lambda) \triangleq \mathcal{L}(g_\Lambda)$. Let $\widehat{\Lambda}_{\mathrm{DCMOML}}$ be any empirical risk minimizer:*

$$\widehat{\Lambda}_{\mathrm{DCMOML}} \in \arg\min_{\Lambda\in\mathcal{H}} \frac{1}{N}\sum_{i=1}^{N}\left(D_{i\kappa_i} - P_{i\kappa_i}^{\top} g_\Lambda(X_i^{-A_i})\right)^2.$$

Assume realizability: there exists $\Lambda_0 \in \mathcal{H}$ such that $g_{\Lambda_0}(X_i^{-A_i}) = g^*(X_i^{-A_i})$ almost surely. Assume further that the induced squared-loss class is Glivenko–Cantelli:

$$\sup_{\Lambda\in\mathcal{H}}\left|\frac{1}{N}\sum_{i=1}^{N}\left(D_{i\kappa_i} - P_{i\kappa_i}^{\top} g_\Lambda(X_i^{-A_i})\right)^2 - \mathcal{L}(\Lambda)\right| \xrightarrow{p} 0.$$

Then $\mathcal{L}(g_{\widehat{\Lambda}_{\mathrm{DCMOML}}}) \xrightarrow{p} \mathcal{L}(g^*)$. Moreover, if X° is the masked input of a new task drawn independently of the training data, then $g_{\widehat{\Lambda}_{\mathrm{DCMOML}}}(X^\circ) \xrightarrow{p} g^*(X^\circ)$.

Remark 2 (Scope of identification)**.** DCMOML identifies $\mathbb{E}[\Theta_i \mid X_i^{-A_i}]$ – the posterior mean of the structural demand parameter given task covariates, realized price history, and unobfuscated outcomes. Four positioning notes follow. First, the target is neither Θ_i (not identifiable from few within-task observations) nor $g(Z_i) = \mathbb{E}[\Theta_i \mid Z_i]$, which discards price-path information. Second, the target is *policy-conditional*: because prices are endogenous, the estimand shifts if the pricing policy changes. Third, identification requires no excluded instruments or randomized prices; it instead relies on availability of at least two locally exogenous prices at the task-level (Assumption 1). Fourth, the target has an explicit observable formula, (16) in Appendix B, so identification does not hinge on the squared-loss characterization or on a parameterization.

The next result shows that the proposed information design is maximally uniformly valid for the squared-loss objective under the two-point query randomization scheme: for every nonconstant refinement that reveals a function of the masked outcomes (D_{ia_i}, D_{ib_i}) or the query index κ_i, there is an admissible data-generating process under which the refined population risk fails to recover the refined conditional target uniquely.

Theorem 2 (Maximal uniform validity)**.** *Fix a two-index selection rule $\mathcal{A}(z, p_{1:K})$, and suppose there exists a support point $(\bar{z}, \bar{p}_{1:K})$ such that $\mathcal{A}(\bar{z}, \bar{p}_{1:K}) = \{a, b\}$, and $\bar{p}_a \neq \bar{p}_b$. Given a data-generating process, let $A_i = \{a_i, b_i\} = \mathcal{A}(Z_i, p_{i1:K})$ and let $X_i^{-A_i} = (Z_i, p_{i1:K}, D_i^{-A_i})$ be the masked-outcome information set. Let κ_i be drawn uniformly from A_i, independently across tasks and independently of $(Z_i, p_{i1:K}, \Theta_i, \epsilon_{i1:K})$ conditional on A_i. Consider any refinement*

$$W_i = (X_i^{-A_i}, R_i), \qquad R_i = r_i(\kappa_i, D_{ia_i}, D_{ib_i}),$$

where the refinement rule is measurable and its restriction to $\{a, b\} \times \mathbb{R}^2$ at the support point $(\bar{z}, \bar{p}_{1:K})$ is nonconstant. Then there exists a data-generating process satisfying the model assumptions, including condition (4) *for every $k \in \{1, \ldots, K\}$, such that the refined population risk*

$$L_W(g) = \mathbb{E}\left[\left(D_{i\kappa_i} - P_{i\kappa_i}^{\top} g(W_i)\right)^2\right]$$

does not identify $\mathbb{E}[\Theta_i \mid W_i]$.

Two failure modes drive this result. If the refinement reveals the query index, the loss constrains only $P_{i\kappa_i}^{\top}\Theta_i$ at that input and leaves the orthogonal direction free. Otherwise it is a nonconstant function of the masked outcomes, and conditioning on it can select on the query shock, so the conditional target is no longer a minimizer. The theorem is a worst-case statement about the specified population loss over the model class: it does not say that every refined target is statistically unidentified (if the refinement reveals only κ_i, the target $\mathbb{E}[\Theta_i \mid X_i^{-A_i}, \kappa_i] = \mathbb{E}[\Theta_i \mid X_i^{-A_i}]$ remains identified by Theorem 1; it is the refined loss that cannot select it), nor that richer information is harmful under stronger assumptions.

6 Experiments

6.1 Evaluating Design Alternatives

We evaluate DCMOML and design alternatives on a more realistic variant of the illustrative example considered in Appendix A. Each task i is a store-specific linear demand curve with task-varying slope

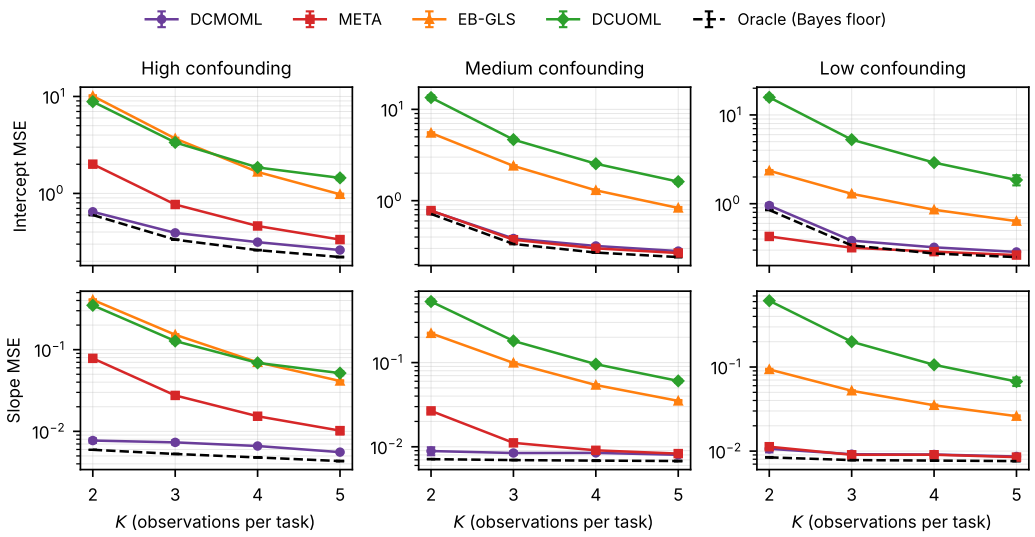


Figure 1: Synthetic study: intercept (top) and slope (bottom) MSE against the number of observations per task, under high, medium and low confounding (log scale; mean over 10 seeds with 95% intervals, mostly smaller than the markers). The oracle is the Bayes floor for DCMOML's information set.

and intercept. Prices are chosen endogenously by a manager who (i) forms a noisy estimate of the task's revenue-optimal price and (ii) experiments locally around that estimate. We vary the noise in this estimate to control confounding: accurate estimates tightly couple prices to latent parameters, while higher noise introduces quasi-exogenous price variation and weakens confounding. We consider high, medium, and low confounding regimes (HC, MC, LC, resp.), always with small within-task experimentation, and $K \in \{2, 3, 4, 5\}$ observations per task. Since prices do not depend on the demand shocks, Assumption 1 is satisfied for any pair of indices; DCMOML queries the lowest- and highest-priced observations of each task. The full setup appears in Appendix C.

Methods. The alternatives isolate key design choices in DCMOML.

1. META. A classical meta-learning baseline that conditions on $(p_1, D_1), \ldots, (p_{K-1}, D_{K-1})$ and minimizes squared error on the final point (p_K, D_K).
2. DCML (Decision-conditioned meta-learning). META that also conditions on the query price p_K (but not D_K). This baseline tests whether simply passing p_K to META resolves the confounding problem.
3. DCUOML (Decision-conditioned unassigned-outcome meta-learning). A DCMOML variant that also reveals the non-query pair outcome to the learner, without assigning it to a price.
4. EB-GLS (Empirical Bayes with GLS shrinkage). A hierarchical random-effects baseline that models (θ_i^0, θ_i^1) as draws from a shared Gaussian prior estimated from the data, and returns their posterior mean under a Gaussian likelihood with the known noise variance.
5. SHARED (Pooled OLS). A single linear demand model fitted by pooling all tasks and observations.
6. TASK-OLS (Per-task OLS). An independent OLS fit for each task using only its own observations.

We also report a reference, *Oracle*: the Bayes floor $\mathbb{E}[\Theta_i \mid X_i^{-A_i}]$ for DCMOML's information set, computed by numerical integration under the known data-generating process. It is the target that Theorem 1 identifies.

All four meta-learners (DCMOML, META, DCML, DCUOML) use the same function class, a polynomial map from the method's information set to (θ_i^0, θ_i^1) that is symmetric in exchangeable observations (Appendix C), fitted exactly by least squares. There is thus no optimization error, and differences between methods reflect only the information design and supervision. For each configuration we train on 20,000 tasks, evaluate on 10,000 new tasks, and repeat over 10 seeds. Figure 1 reports MSE for the slope θ_i^1 and intercept θ_i^0; SHARED, DCML and TASK-OLS are far off the plotted scale and appear, with all other results, in Tables 2 and 3 (Appendix C).

Results. We make four main observations. (1) Under high confounding, DCMOML is the most accurate method at every K, for both parameters and in every seed, and stays close to its Bayes floor: intercept MSE 0.65 against a floor of 0.60 at $K = 2$ (META 2.00, EB-GLS 10.1, SHARED 25.8), and 0.26 against 0.22 at $K = 5$ (META 0.33). DCML fails at every K (intercept MSE about 96) because once the query price is an input the loss no longer pins down the coefficient vector; with the polynomial learner the failure is literal, as its least-squares design is rank-deficient. EB-GLS improves on per-task estimation through shrinkage, but its prior treats parameters as independent of prices, so it remains well behind DCMOML whenever confounding is present. (2) As confounding weakens, the gap between DCMOML and META closes. Under medium confounding the two are within about 0.02 in intercept MSE for $K \geq 3$, with DCMOML still clearly better on the slope at $K \leq 3$. Under low confounding META has the lower intercept MSE at every K (0.43 versus 0.95 at $K = 2$; 0.26 versus 0.29 at $K = 5$), with similar slope errors: masking discards the demand observations META conditions on, and when prices carry little confounding that information outweighs META's bias. META can then even beat DCMOML's oracle, which bounds only methods restricted to DCMOML's information set. (3) DCUOML performs poorly at every K despite query randomization: revealing the non-query outcome lets the learner partly infer the query decision, reintroducing the same failure. This is consistent with Theorem 2, which guarantees such a failure for some process in the model class; the experiment shows it also occurs in this one. (4) Unmasked outcomes make the problem easier: under high confounding the oracle intercept MSE falls from 0.60 at $K = 2$ to 0.22 at $K = 5$, and DCMOML's error falls with it. Theorem 1 implies that the population risk is minimized at the oracle target, so the remaining gap reflects the learner's finite-sample and approximation error. Consistent with this, the gap narrows as tasks are added (Table 4): at $K = 2$, with 400,000 tasks, DCMOML reaches its floor (0.601 versus 0.600) while META stays near 2.0, and at $K = 5$ moving from 20,000 to 50,000 tasks lowers its intercept MSE from 0.26 to 0.24 (floor 0.22).

6.2 Evaluation on Retail Dataset

We evaluate DCMOML on *UK-online-retail*, an online retail transaction dataset from a UK-based gift retailer spanning 01/12/2010–09/12/2011 [Chen, 2015]. The processed panel records, for each product and day with at least one sale, the transaction price and the number of units sold. It covers 3,806 products with an average of 3.79 distinct prices per product (median 4). Since ground-truth demand parameters Θ_i are unobserved, this experiment cannot test causal recovery; instead, it asks whether the DCMOML information design remains useful for predicting demand at a pricing decision withheld from each product's training data.

Tasks and holdout protocol. We define two product-level tasks. In both, the context is $Z_i \in \mathbb{R}^{1024}$, the sentence-transformer embedding [Reimers and Gurevych, 2019] of the product title, and the transfer-learning methods are trained on two observations per product ($K = 2$), with a third observation held out.

1. *Static-Top3 (withheld price).* We retain products with at least three distinct prices ($n = 2{,}833$). For each product, we compute average daily demand over the sale days at each of its three most frequently observed prices, producing pairs $(p_{i1}, D_{i1}), (p_{i2}, D_{i2}), (p_{i3}, D_{i3})$ ordered by frequency. We train on the top two pairs and evaluate on the third. The held-out price is never seen for that product during training, and it lies outside the interval spanned by the two training prices for 86% of products, so this task tests prediction at a less frequently used, mostly out-of-range price.
2. *Exposure-Sequence (later pricing episode).* For each product, we compress the daily series into a sequence of price exposures $\{(p_{ik}, e_{ik}, D_{ik})\}_k$, where p_{ik} is the kth distinct price run, e_{ik} is the number of consecutive sale days in that run, and D_{ik} is average daily demand over those days. By construction $p_{i,k} \neq p_{i,k+1}$. We retain products with at least three exposures ($n = 3{,}132$), train on the first two, and evaluate on the third. The held-out price may coincide with a training price, but it is observed in a later episode under a different history.

Methods and protocol. We compare (i) DCMOML, with inputs (Z_i, p_{i1}, p_{i2}) for Static-Top3 and $(Z_i, p_{i1}, e_{i1}, p_{i2}, e_{i2})$ for Exposure-Sequence, trained with the loss averaged over query indices $\{1, 2\}$; (ii) META, which conditions on one support pair and predicts the other, symmetrized over both assignments; (iii) SHARED, which predicts Θ_i from Z_i alone; and (iv) PER-TASK, a separate exposure-weighted linear fit per product on all non-holdout prices (Static-Top3 only). Since D_{ik} averages over a variable number of days, all losses weight observations by exposure, normalized

Table 1: UK-online-retail, 100 seeds: mean $\pm$ 95% CI; paired rows give the mean DCMOML$-$baseline difference and the share of seeds in which DCMOML is lower.

Task	Metric	DCMOML	META	SHARED
Static-Top3	Held-out RMSE	**109.45** $\pm$ 0.06	110.08 $\pm$ 0.24	110.13 $\pm$ 0.10
	paired diff. (share)	–	-0.63 ± 0.25 (68%)	-0.68 ± 0.09 (93%)
Exposure-Sequence	Held-out RMSE	**35.08** $\pm$ 0.41	39.26 $\pm$ 0.75	35.63 $\pm$ 0.25
	paired diff. (share)	–	-4.19 ± 0.54 (98%)	-0.56 ± 0.37 (80%)

within product; all transfer methods use the same two-hidden-layer MLP (Appendix D). For each of 100 seeds, products are split 80/20; observations 1–2 of the 80% subset are used for fitting and of the 20% subset for early stopping, and observation 3 is never used for either. We report the held-out RMSE of the predicted demand $\widehat{\theta}_i^0 + \widehat{\theta}_i^1 p_{i3}$ on observation 3 of all eligible products, exposure-weighted; all methods share the split in each seed, so we also report paired differences across seeds.

Findings. Table 1 summarizes the results. On Static-Top3, DCMOML has the lowest held-out RMSE, below META in 68% of seeds and below SHARED in 93%. The mean gap to META is driven partly by a few seeds in which META fits poorly; the median paired difference is -0.16. PER-TASK performs poorly (RMSE 200.50), highlighting the gains from cross-task transfer despite product heterogeneity.

On Exposure-Sequence, DCMOML again has the lowest held-out RMSE, below META in 98% of seeds and below SHARED in 80%. The fitted slopes, however, are not interpretable as price responses: in the median seed only 39% of DCMOML's slopes are negative, and this share ranges from 0% to 100% across seeds (80% for META, 67% for SHARED). Exposures are short and volatile (62% of training exposures last a single sale day), so the training pairs carry little within-product price information. We therefore read this task as evidence that the information design improves prediction in a later pricing episode, not as evidence about price response.

Two caveats apply to both tasks: neither verifies Assumption 1 or recovers causal parameters, and transaction prices in this dataset may vary with order size and customer type, with days without sales unrecorded. The retail results are predictive evidence that complements, rather than replicates, the synthetic causal validation of Section 6.1.

7 Conclusion

We show how learning across tasks can improve demand estimates even when historical prices reflect unobserved differences in demand determinants. Our approach uses information contained in those pricing decisions while carefully controlling which outcomes enter the learner's inputs and supervision.

A limitation of our guarantees is that they rely on linear demand and the availability of two distinct locally exogenous prices within each task. Local exogeneity is determined by the price-setting process and cannot be verified from the observed data alone. For example, an adaptive pricing policy can violate this assumption when later prices respond to demand shocks at the selected observations. Even when the local exogeneity assumption holds, closely spaced prices contain less information about the demand slope, and learning a flexible mapping from task histories to demand parameters requires many tasks.

As we note, there are also limits to the benefits of the information design. Masking outcomes sacrifices information, and our experiments show that this cost can outweigh the benefit when confounding is weak. Moreover, the conditional-mean target reflects the historical pricing policy, so a change in that policy may require refitting the learned mapping. The retail experiments demonstrate that the design can improve demand prediction in practice; however, we do not claim that the resulting estimates recover causal price responses.

These limitations suggest several directions for future work: extending the approach to nonlinear demand and cross-price effects, assessing sensitivity to violations of local exogeneity, and investigating practical guidelines for deciding between DCMOML and standard meta-learning approaches.

Acknowledgments and Disclosure of Funding

This work was supported by the Discovery Partners Institute (DPI) of the University of Illinois System through a DPI Science Team seed grant (Award No. 11575), "Industry Data Trusts using Next-Generation AI Technologies." We also thank the anonymous referees and the area chair for their helpful comments and feedback. The authors declare no competing interests.

References

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

Susan Athey and Guido W. Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

Jonathan Baxter. A model of inductive bias learning. *Journal of Artificial Intelligence Research*, 12: 149–198, 2000.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890, 1995.

Steven T. Berry. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262, 1994.

Steven T. Berry and Philip A. Haile. Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797, 2014.

Rich Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, 1997.

Gary Chamberlain. Multivariate regression models for panel data. *Journal of Econometrics*, 18(1): 5–46, 1982.

Gary Chamberlain. Panel data. In *Handbook of Econometrics*, volume 2, pages 1247–1318. Elsevier, 1984.

Daqing Chen. Online Retail. UCI Machine Learning Repository, 2015. DOI: https://doi.org/10.24432/C5BW33.

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.

Bradley Efron. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press, 2010. ISBN 9780521192491. doi: 10.1017/CBO9780511761362.

Theodoros Evgeniou and Massimiliano Pontil. Regularized multi-task learning. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 109–117, 2004.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 1126–1135, 2017.

Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. CRC Press, third edition, 2013.

Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. Deep IV: A flexible approach for counterfactual prediction. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1414–1423. PMLR, 2017.

Jerry A. Hausman. Valuation of new goods under perfect and imperfect competition. In Timothy F. Bresnahan and Robert J. Gordon, editors, *The Economics of New Goods*, pages 207–248. University of Chicago Press, 1996.

Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9): 5149–5169, 2022. doi: 10.1109/TPAMI.2021.3079209. Also available as arXiv:2004.05439.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

Jack Kiefer and Jacob Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466, 1952. doi: 10.1214/aoms/1177729392.

Nan M. Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4): 963–974, 1982.

Yair Mundlak. On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85, 1978.

Aviv Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2): 307–342, 2001.

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

Amil Petrin and Kenneth Train. A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1):3–13, 2010.

Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, 2019.

Herbert Robbins. An empirical Bayes approach to statistics. In Jerzy Neyman, editor, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 157–163. University of California Press, Berkeley, CA, 1956.

Sentence-Transformers. all-roberta-large-v1. Hugging Face model card, 2021. `https://huggingface.co/sentence-transformers/all-roberta-large-v1`.

P. A. V. B. Swamy. Efficient inference in a random coefficient regression model. *Econometrica*, 38 (2):311–323, 1970.

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.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, second edition, 2010.

A Illustrative example

Shared-Model Estimation Can Fail under Near-Optimal Local Pricing behavior. We illustrate the failure of shared-model learning with a simple example in which learning only the shared component leads to a qualitatively incorrect conclusion. Suppose all stores share the same observable context, $Z_i \equiv Z$ for all i, so that $g(Z_i)$ is constant across tasks. In this case, the estimator in (6) reduces to fitting a *single* common parameter vector Θ using all pooled observations. Assume that each store i has a true linear demand curve

$$D_{ik} = \theta_i^0 + \theta_1 p_{ik} + \epsilon_{ik}, \qquad \theta_1 < 0,$$

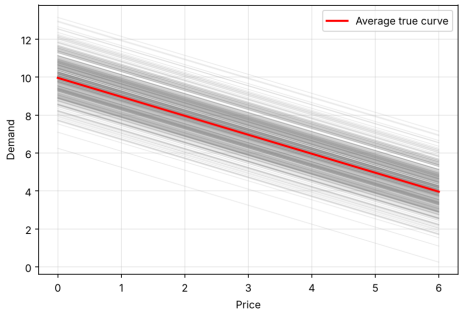
where θ_1 is a common (negative) slope and $\theta_i^0 \geq 0$ is a store-specific intercept. Suppose further that the store manager observes her local demand curve perfectly and sets prices close to the revenue-maximizing level

$$p_i^* = \arg\max_p \; p(\theta_i^0 + \theta_1 p) = -\frac{\theta_i^0}{2\theta_1},$$

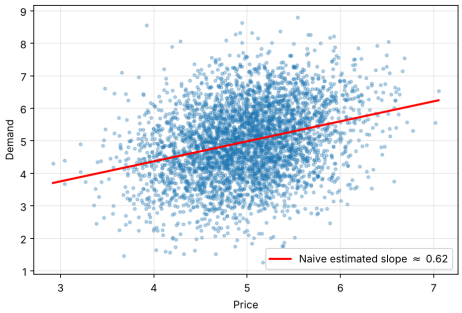
with small idiosyncratic pricing noise: $p_{ik} = p_i^* + \varepsilon_{ik}$. Substituting into the demand equation yields

$$D_{ik} = \theta_i^0 + \theta_1 p_{ik} + \epsilon_{ik} = \frac{\theta_i^0}{2} + \theta_1 \varepsilon_{ik} + \epsilon_{ik} = -\theta_1 p_{ik} + 2\theta_1 \varepsilon_{ik} + \epsilon_{ik}.$$

Thus, in the pooled data, demand is positively correlated with price even though the true causal slope θ_1 is strictly negative. Because stores with higher intercepts optimally charge higher prices and also exhibit higher demand, the shared-model estimator attributes this cross-store variation to a positive price effect. As a result, the estimator in (6) converges to a parameter vector with a positive price coefficient, reversing the sign of the true causal effect. Figure 2 illustrates this phenomenon: although all stores have downward-sloping demand curves, the pooled price–demand relationship is upward sloping due to endogenous pricing.



(a) True demand curves across stores. Each store has a linear mean demand curve $D_{ik} = \theta_i^0 + \theta_1 p_{ik}$ with heterogeneous intercepts $\theta_i^0 \sim \mathcal{N}(10, 1)$ and identical negative slope $\theta_1 = -1$.



(b) Pooled price–demand data. Managers set prices near the optimum $p_i^* = -\theta_i^0/(2\theta_1)$. Demand noise is $\epsilon_{ik} \sim \mathcal{N}(0, 1)$, and pricing noise is $\varepsilon_{ik} \sim \mathcal{N}(0, 0.25^2)$.

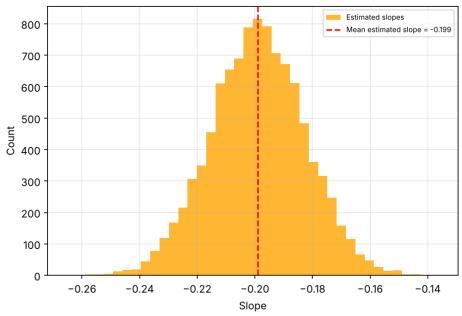
Figure 2: Illustration of confounding in the multi-task pricing setting ($N = 2000$, $K = 2$). Left: True (causal) demand curves all have negative slopes. Right: Pooled data exhibit a positive price–demand relationship due to optimal pricing responding to store-specific intercepts.

Failure of Meta-Learning. We continue with the same example as above with $K = 2$ observations per store. The support set consists of a single observation (p_{i1}, D_{i1}), and the query observation is (p_{i2}, D_{i2}). Suppose the intercept θ_i^0 is Gaussian, and both the demand noise ϵ_{ik} and the pricing noise ε_{ik} are Gaussian. Under the manager's pricing rule, the random vector $\big(\Theta_i, p_{i1}, D_{i1}, p_{i2}, D_{i2}\big)$ is jointly Gaussian. Consequently, the conditional expectation $\mathbb{E}[\Theta_i \mid p_{i1}, D_{i1}]$ is an affine function of (p_{i1}, D_{i1}), and hence, linear adaptation rules of the form $\widehat{\Theta}_i = A[p_{i1}, D_{i1}]^\top + B$ are sufficient to recover the mapping. Nevertheless, the ERM objective (7) remains biased since the second-period price p_{i2} is endogenously chosen and remains correlated with Θ_i even after conditioning on (p_{i1}, D_{i1}). Thus, meta-learning fails to recover the causal price-response function. Figure 3 illustrates this failure.

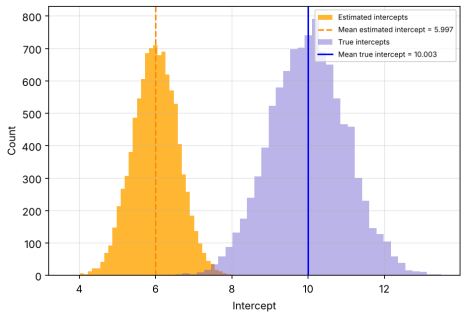
Performance of DCMOML. We now examine the performance of DCMOML in this example, when each store provides only $K = 2$ observations. Since the observable context is constant (so $Z_i \equiv Z$), the two-outcome withholding design in (11) leaves the learner with only the realized prices: $X_i^{-A_i} = (p_{i1}, p_{i2})$, with $A_i = \{1, 2\}$ for every task. Thus, DCMOML amounts to training a meta-learner that maps the *pair of prices* to the store-level demand parameters,

$$g^*(p_{i1}, p_{i2}) = \mathbb{E}[\Theta_i \mid p_{i1}, p_{i2}], \qquad \Theta_i = (\theta_i^0, \theta_i^1)^\top.$$

As discussed earlier, in this example, all primitives are Gaussian, and the pricing rule is affine, so $(\Theta_i, p_{i1}, p_{i2})$ is jointly Gaussian. Consequently, $g^*(p_{i1}, p_{i2})$ is an affine function of (p_{i1}, p_{i2}), and it is sufficient to fit a linear predictor. Moreover, (p_{i1}, p_{i2}) is exchangeable, so the target is symmetric:



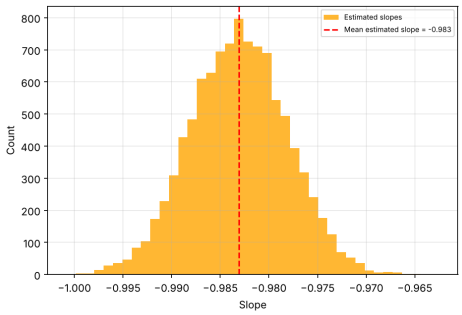
(a) Distribution of estimated slopes. The true slope is -1 for all tasks.



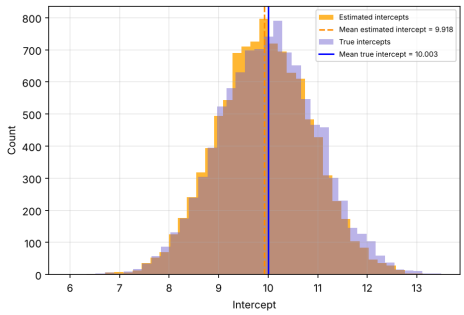
(b) Distribution of true and estimated intercepts.

Figure 3: Estimation performance of the simple outcome-based meta-estimator $\hat{\theta}_{ij} = a_j p_{i1} + b_j D_{i1} + c_j$ for $j \in \{0, 1\}$ with $K = 2$, fitted on 100,000 stores and evaluated on 10,000 new stores. The left panel shows slope estimates while the right panel shows intercept estimates.

$g^*(p_{i1}, p_{i2}) = g^*(p_{i2}, p_{i1})$. One can therefore impose a symmetric linear form in which the two prices enter with the same coefficients. Figure 4 shows the resulting estimates of the conditional mean demand parameters given the two observed prices.



(a) Distribution of estimated slopes under DCMOML. The true slope is -1 for all tasks.



(b) Distribution of true and estimated intercepts under DCMOML.

Figure 4: Estimation performance of the DCMOML meta-learner $\hat{\theta}_{ij} = a_j(p_{i1} + p_{i2}) + c_j$ for $j \in \{0, 1\}$ with $K = 2$, fitted on 100,000 stores and evaluated on 10,000 new stores.

B Proofs of Main Results

Conventions. Population expectations involving a fitted predictor are taken over an independent draw from the task population, conditional on the training data $\mathcal{T}_N$. Fitted estimators and the uniform deviations in Theorem 1 are assumed measurable. Refinement statistics in Theorem 2 take values in a standard Borel space, as ordinary real- or vector-valued features do, so that distinct values are measurably distinguishable.

Proof of Theorem 1. Suppress the task index. Write $X = X^{-A}$ and $A = \{a, b\}$, and abbreviate the query regressor as $P = P_\kappa$. Because A is a measurable function of $(Z, p_{1:K})$, which is part of X, the pair A and its prices p_a, p_b are X-measurable. By Assumption 2, Θ and each $P_k^\top \Theta$ are square integrable, and each shock ϵ_k has finite variance by the model, so each $D_k = P_k^\top \Theta + \epsilon_k$ is square integrable.

Step 1: conditional regression identity. We first show that the two masked shocks have mean zero given the input:

$$\mathbb{E}[\epsilon_a \mid X] = \mathbb{E}[\epsilon_b \mid X] = 0. \tag{13}$$

The labels a, b are random, so we argue separately on each of the finitely many events $E_{k\ell} = \{a = k, b = \ell\}$, which are $(Z, p_{1:K})$-measurable. On $E_{k\ell}$ the input equals $X_{k\ell} = (Z, p_{1:K}, \{D_j :$

$j \notin \{k, \ell\}\}$), which is a function of $(Z, p_{1:K}, \mathbf{D}^{-k})$. Hence, for every bounded measurable h, Assumption 1 for index k on $E_{k\ell}$ gives

$$\mathbb{E}\big[\epsilon_k \mathbf{1}_{E_{k\ell}} h(X)\big] = \mathbb{E}\big[\mathbf{1}_{E_{k\ell}} h(X_{k\ell})\, \mathbb{E}[\epsilon_k \mid Z, p_{1:K}, \mathbf{D}^{-k}]\big] = 0,$$

that is, $\mathbb{E}[\epsilon_k \mathbf{1}_{E_{k\ell}} \mid X] = 0$. The same argument with ℓ in place of k gives $\mathbb{E}[\epsilon_\ell \mathbf{1}_{E_{k\ell}} \mid X] = 0$. Summing over the events gives (13).

Next, the query. Given A, the draw κ is uniform on A and independent of $(Z, p_{1:K}, \Theta, \epsilon_{1:K})$. Since A and X are functions of these variables, κ is also uniform on A and independent of $(\Theta, \epsilon_{1:K})$ given X. Therefore

$$\begin{aligned}
\mathbb{E}[\Theta \mid X, \kappa] &= \mathbb{E}[\Theta \mid X] \triangleq g^*(X), \\
\mathbb{E}[\epsilon_\kappa \mid X, \kappa] &= \mathbf{1}\{\kappa = a\}\, \mathbb{E}[\epsilon_a \mid X] + \mathbf{1}\{\kappa = b\}\, \mathbb{E}[\epsilon_b \mid X] = 0,
\end{aligned}$$

the last equality by (13). Finally, $P = (1, p_\kappa)^\top$ is (X, κ)-measurable. Taking conditional expectations in $D_\kappa = P^\top \Theta + \epsilon_\kappa$ and pulling P out of the conditional expectation gives

$$\mathbb{E}[D_\kappa \mid X, \kappa] = P^\top g^*(X). \tag{14}$$

Pulling P out is valid since θ^0 and $\theta^0 + p_\kappa \theta^1 = P^\top \Theta$ are integrable by Assumption 2, hence so is $p_\kappa \theta^1$.

Step 2: excess-risk identity and uniqueness. Conditional Jensen's inequality applied to (14) gives $\mathbb{E}[(P^\top g^*(X))^2] \le \mathbb{E}[D_\kappa^2] < \infty$, so g^* satisfies the integrability requirement of the theorem and has finite risk. Fix any other g that satisfies it, put $\Delta(X) = g(X) - g^*(X)$, and let $\xi = D_\kappa - P^\top g^*(X)$ be the residual of the target. Both ξ and $P^\top \Delta(X)$ are square integrable, and $\mathbb{E}[\xi \mid X, \kappa] = 0$ by (14), so

$$\mathbb{E}\big[\xi\, P^\top \Delta(X)\big] = \mathbb{E}\big[P^\top \Delta(X)\, \mathbb{E}[\xi \mid X, \kappa]\big] = 0.$$

Since $D_\kappa - P^\top g(X) = \xi - P^\top \Delta(X)$, expanding the square gives

$$\mathcal{L}(g) - \mathcal{L}(g^*) = \mathbb{E}\big[(P^\top \Delta(X))^2\big] = \mathbb{E}\big[\Delta(X)^\top Q(X) \Delta(X)\big], \tag{15}$$

where $Q(X) = \frac{1}{2} P_a P_a^\top + \frac{1}{2} P_b P_b^\top$; the second equality conditions on X and averages over the two equally likely values of κ. For every $v \in \mathbb{R}^2$, $v^\top Q(X) v = \frac{1}{2}(P_a^\top v)^2 + \frac{1}{2}(P_b^\top v)^2$; because $p_a \neq p_b$ almost surely, P_a and P_b are linearly independent and $Q(X)$ is positive definite almost surely. The integrand $\Delta^\top Q \Delta$ is therefore nonnegative and vanishes only where $\Delta(X) = 0$, so $\mathcal{L}(g) = \mathcal{L}(g^*)$ forces $\Delta(X) = 0$ almost surely. Thus g^* is the unique minimizer up to almost-sure equality. Since $\mathcal{L}$ depends only on the law of the observed data (X, D_a, D_b) and on the known query randomization, the target is identified.

Step 3: consistency. Let $\widehat{\mathcal{L}}_N$ be the empirical risk in the theorem and $\delta_N = \sup_{\Lambda \in \mathcal{H}} |\widehat{\mathcal{L}}_N(\Lambda) - \mathcal{L}(\Lambda)|$, so $\delta_N \to 0$ in probability by the Glivenko–Cantelli assumption. Let $\widehat{\Lambda}$ be an empirical risk minimizer and Λ_0 the parameter in the realizability assumption, so that $\mathcal{L}(\Lambda_0) = \mathcal{L}(g^*)$. On the event $\{\delta_N < \infty\}$, whose probability tends to one,

$$\mathcal{L}(\widehat{\Lambda}) \le \widehat{\mathcal{L}}_N(\widehat{\Lambda}) + \delta_N \le \widehat{\mathcal{L}}_N(\Lambda_0) + \delta_N \le \mathcal{L}(g^*) + 2\delta_N,$$

using the definition of δ_N, then the optimality of $\widehat{\Lambda}$ for the empirical risk, then the definition of δ_N again. Together with $\mathcal{L}(\widehat{\Lambda}) \ge \mathcal{L}(g^*)$ from Step 2, this gives $0 \le \mathcal{L}(\widehat{\Lambda}) - \mathcal{L}(g^*) \le 2\delta_N \to 0$ in probability, which is risk consistency.

For coefficient consistency, let $\mathcal{T}_N$ be the training data (including the query draws), let X° be the masked input of a new task drawn independently of $\mathcal{T}_N$, and write $\Delta_N(x) = g_{\widehat{\Lambda}}(x) - g^*(x)$ and $r_N = \mathcal{L}(\widehat{\Lambda}) - \mathcal{L}(g^*)$. On $\{\delta_N < \infty\}$ the fitted function has finite risk, and hence satisfies the integrability requirement, since $(P^\top g)^2 \le 2D_\kappa^2 + 2(D_\kappa - P^\top g)^2$. Applying (15) to it with $\mathcal{T}_N$ held fixed gives

$$\mathbb{E}\big[\Delta_N(X^\circ)^\top Q(X^\circ) \Delta_N(X^\circ) \,\big|\, \mathcal{T}_N\big] = r_N.$$

Let $\lambda(x) = \lambda_{\min}(Q(x))$, which is positive almost surely by Step 2 but need not be bounded away from zero. Fix $\varepsilon > 0$ and $t > 0$. If $\|\Delta_N(X^\circ)\|_2 > \varepsilon$, then either $\lambda(X^\circ) \le t$, or $\lambda(X^\circ) > t$ and

hence $\Delta_N^\top Q \Delta_N \geq \lambda \|\Delta_N\|_2^2 > t\varepsilon^2$ at X°. Bounding the probability of the second event by Markov's inequality gives

$$\mathbb{P}\big(\|\Delta_N(X^\circ)\|_2 > \varepsilon \,\big|\, \mathcal{T}_N\big) \leq \mathbb{P}\big(\lambda(X^\circ) \leq t\big) + \min\Big\{1, \frac{r_N}{t\varepsilon^2}\Big\}.$$

Since $r_N \to 0$ in probability, taking expectations over $\mathcal{T}_N$ and using bounded convergence gives $\limsup_N \mathbb{P}(\|\Delta_N(X^\circ)\|_2 > \varepsilon) \leq \mathbb{P}(\lambda(X^\circ) \leq t)$ for every $t > 0$. Letting $t \downarrow 0$, the right side tends to zero because $\lambda(X^\circ) > 0$ almost surely. Hence $g_{\widehat{\Lambda}}(X^\circ) \to g^*(X^\circ)$ in probability. □

Remarks on Theorem 1. Two consequences of the proof are worth recording.

(i) An observable formula for the target. Let $m_k(X) = \mathbb{E}[D_k \mid X]$ for $k \in \{a, b\}$. Since p_k is X-measurable, (13) gives $m_k = P_k^\top g^*$, and solving these two linear equations gives

$$g^*(X) = \Big(\frac{p_b\, m_a(X) - p_a\, m_b(X)}{p_b - p_a},\ \frac{m_b(X) - m_a(X)}{p_b - p_a}\Big)^\top. \tag{16}$$

Every term on the right is determined by the law of the observed data, so this formula establishes identification directly, without reference to the squared-loss risk.

(ii) Approximate minimization and misspecification. Exact realizability and exact optimization are not needed for the risk bound. Suppose $\widehat{g}_N$ satisfies $\widehat{\mathcal{L}}_N(\widehat{g}_N) \leq \inf_{g \in \mathcal{H}_N} \widehat{\mathcal{L}}_N(g) + \eta_N$ over a possibly growing class $\mathcal{H}_N$, and let δ_N be the uniform deviation over $\mathcal{H}_N$. The argument of Step 3, combined with (15), gives the oracle inequality

$$0 \leq \mathcal{L}(\widehat{g}_N) - \mathcal{L}(g^*) \leq \inf_{g \in \mathcal{H}_N} \mathbb{E}\big[(g - g^*)^\top Q\,(g - g^*)\big] + 2\delta_N + \eta_N,$$

so risk consistency follows whenever the approximation term, δ_N and η_N all vanish.

Proof of Theorem 2. Suppress the task index. Fix the support point $(\bar{z}, \bar{p}_{1:K})$ from the statement, and write $r(k, u, v)$ for the refinement rule at that point, a measurable function on $\{a, b\} \times \mathbb{R}^2$ that is nonconstant by assumption. Write $P_j = (1, \bar{p}_j)^\top$; since $\bar{p}_a \neq \bar{p}_b$, the vectors P_a and P_b are linearly independent.

In each of the two cases below we construct a process with the following common features: $Z \equiv \bar{z}$ and $p_{1:K} \equiv \bar{p}_{1:K}$, so that $A = \{a, b\}$ always; $\Theta \equiv \bar{\theta}$ is deterministic (the shared component equals $\bar{\theta}$ at $\bar{z}$ and $\Omega \equiv 0$); the shocks $\epsilon_1, \ldots, \epsilon_K$ are i.i.d., bounded and mean zero; κ is uniform on $\{a, b\}$ and independent of everything else; and tasks are i.i.d. Because prices and coefficients are deterministic, $\mathbf{D}^{-j}$ is a function of the shocks other than ϵ_j, which are independent of ϵ_j, so $\mathbb{E}[\epsilon_j \mid Z, p_{1:K}, \mathbf{D}^{-j}] = \mathbb{E}[\epsilon_j] = 0$ for every $j = 1, \ldots, K$. Thus every model assumption holds, including condition (4) at every index and Assumption 2. Because Θ is deterministic, the refined target is $\mathbb{E}[\Theta \mid W] = \bar{\theta}$. It remains to show that the refined risk either has a minimizer other than $\bar{\theta}$ or is not minimized by $\bar{\theta}$.

Case 1: the refinement reveals the query index. Suppose $r(a, u, v) \neq r(b, u, v)$ for some $(u, v) \in \mathbb{R}^2$. Let $\bar{\theta}$ be the unique vector with $P_a^\top \bar{\theta} = u$ and $P_b^\top \bar{\theta} = v$, and set all shocks to zero. Then $(D_a, D_b) = (u, v)$ always, so $R = r(\kappa, u, v)$ takes different values for $\kappa = a$ and $\kappa = b$, and κ is a measurable function of R, hence of W. Consider

$$\widetilde{g}(W) = \bar{\theta} + (\bar{p}_\kappa, -1)^\top.$$

Because $P_\kappa^\top (\bar{p}_\kappa, -1)^\top = 0$, we have $P_\kappa^\top \widetilde{g}(W) = P_\kappa^\top \bar{\theta} = D_\kappa$. Both $\bar{\theta}$ and $\widetilde{g}$ therefore have zero refined risk, so both minimize L_W, and $\widetilde{g} \neq \bar{\theta}$ everywhere. The refined risk has non-unique minimizers.

Case 2: the refinement does not reveal the query index. Otherwise $r(a, s, t) = r(b, s, t) \triangleq F(s, t)$ for all (s, t), so $R = F(D_a, D_b)$ and F is nonconstant. A function on $\mathbb{R}^2$ that is constant along every horizontal and every vertical line is constant, so F changes along one of its two coordinates. Relabeling a and b if necessary, there exist $u_0 < u_1$ and c with $F(u_0, c) \neq F(u_1, c)$. Put $\mu = (u_0 + u_1)/2$ and $d = (u_1 - u_0)/2 > 0$, and let $\bar{\theta}$ be the unique vector with $P_a^\top \bar{\theta} = \mu$ and $P_b^\top \bar{\theta} = c$. Let the shocks be i.i.d. with

$$\mathbb{P}(\epsilon_j = -d) = \mathbb{P}(\epsilon_j = d) = \tfrac{1}{8}, \qquad \mathbb{P}(\epsilon_j = 0) = \tfrac{3}{4}.$$

Then $D_a = \mu + \epsilon_a$ and $D_b = c + \epsilon_b$. Let $B = \{R = F(u_0, c)\}$, an event determined by W, and let $m = \mathbb{E}[\epsilon_a \mathbf{1}_B]$. We bound m by splitting on the values of (ϵ_a, ϵ_b):

- if $\epsilon_a = 0$, the contribution is zero;
- if $\epsilon_a = -d$ and $\epsilon_b = 0$, then $(D_a, D_b) = (u_0, c)$, so B occurs; this contributes $-d \cdot \frac{1}{8} \cdot \frac{3}{4}$;
- if $\epsilon_a = d$ and $\epsilon_b = 0$, then $(D_a, D_b) = (u_1, c)$, so B does not occur; this contributes zero;
- if $\epsilon_a = -d$ and $\epsilon_b \neq 0$, the contribution is at most zero;
- if $\epsilon_a = d$ and $\epsilon_b \neq 0$, the contribution is at most $d \cdot \frac{1}{8} \cdot \frac{1}{4}$.

Hence $m \leq -\frac{3d}{32} + \frac{d}{32} = -\frac{d}{16} < 0$, and in particular $\mathbb{P}(B) > 0$. Let $v = (-\bar{p}_b, 1)^\top / (\bar{p}_a - \bar{p}_b)$, which satisfies $P_a^\top v = 1$ and $P_b^\top v = 0$, and for $t > 0$ consider the predictor

$$g_t(W) = \bar{\theta} - t\, v\, \mathbf{1}_B.$$

Since $D_\kappa - P_\kappa^\top g_t(W) = \epsilon_\kappa + t\, \mathbf{1}\{\kappa = a\}\mathbf{1}_B$, expanding the square gives

$$L_W(g_t) - L_W(\bar{\theta}) = 2t\, \mathbb{E}\big[\epsilon_\kappa \mathbf{1}\{\kappa = a\}\mathbf{1}_B\big] + t^2\, \mathbb{E}\big[\mathbf{1}\{\kappa = a\}\mathbf{1}_B\big] = t\, m + \tfrac{t^2}{2}\, \mathbb{P}(B),$$

where the last step uses that $R = F(D_a, D_b)$ does not involve κ, so κ is independent of $(\epsilon_a, \mathbf{1}_B)$ and $\mathbb{P}(\kappa = a) = \frac{1}{2}$. Taking $t = d/16$ and using $\mathbb{P}(B) \leq 1$,

$$L_W(g_t) - L_W(\bar{\theta}) \leq -\frac{d^2}{256} + \frac{d^2}{512} = -\frac{d^2}{512} < 0.$$

All risks are finite because prices, coefficients and shocks take finitely many values. The refined target $\bar{\theta} = \mathbb{E}[\Theta \mid W]$ is therefore not a minimizer of the refined risk.

The two cases are exhaustive. In Case 1 the refined risk has distinct minimizers, and in Case 2 it is not minimized by the refined target; in either case it does not identify $\mathbb{E}[\Theta \mid W]$. □

C Experimental setup details for Section 6.1

Data-generating process. Each task i follows the linear demand model

$$D_{ik} \;=\; \theta_i^0 + \theta_i^1\, p_{ik} + \epsilon_{ik}, \qquad k = 1, \ldots, K,$$

with $\theta_i^0 \sim \mathcal{N}(10, 1)$ and $\theta_i^1 \sim \mathcal{N}(-1, 0.1^2)$ truncated to $\theta_i^1 \leq -0.3$ (the truncation keeps optimal prices positive and binds with negligible probability). Demand shocks are i.i.d. $\epsilon_{ik} \sim \mathcal{N}(0, 0.5^2)$, independent of the task parameters and of the manager's pricing noise; the standard deviation is about one tenth of mean demand. Prices are set by a manager who targets a noisy version of the revenue-optimal price $p_i^* = -\theta_i^0/(2\theta_i^1)$:

$$\tilde{p}_i \;=\; p_i^* + \zeta_i, \quad \zeta_i \sim \mathcal{N}\big(0, (\sigma_c |p_i^*|)^2\big), \qquad p_{ik} \;=\; \tilde{p}_i + \nu_{ik}, \quad \nu_{ik} \sim \mathcal{N}\big(0, (0.1|\tilde{p}_i|)^2\big),$$

where $\sigma_c \in \{0, 0.1, 0.2\}$ controls confounding strength (HC, MC, LC). Prices depend on Θ_i but not on the demand shocks, so Assumption 1 holds for every pair of indices. There are no covariates (Z_i is constant).

Query pair and information sets. DCMOML uses the pair $A_i = \{\arg\min_k p_{ik}, \arg\max_k p_{ik}\}$, selected from prices only; for $K = 2$ this is $\{1, 2\}$. Its information set is $X_i^{-A_i} = (p_{i1:K}, \mathbf{D}_i^{-A_i})$. META conditions on observations $1, \ldots, K-1$ and is supervised by observation K. DCML adds p_{iK} to META's inputs. DCUOML adds the non-query pair outcome to DCMOML's inputs without its price index; each training task contributes both query assignments, and predictions average the two.

Learner. All meta-learners (DCMOML, META, DCML, DCUOML) use the same function class, a symmetric polynomial $g(x) = (\phi(x)^\top b_0,\ \phi(x)^\top b_1)$. Observations that play the same role within a task are exchangeable in this DGP, so the learner sees them as a set, through their means and (co)variances: DCMOML's unmasked observations, and META's (and DCML's) support set. Let f collect the first-order inputs: the pair prices for DCMOML, the query price for DCML, the unassigned outcome for DCUOML, and the means of the set. Let s collect the set's variances of prices and demands and their covariance (present when the set has at least two observations). The basis ϕ consists of all monomials of the standardized f up to degree 4, plus each standardized element of s multiplied by all monomials of f up to degree 2. For DCMOML this gives 15 terms per parameter at $K = 2$, 70 at $K = 3$ and 115 at $K \in \{4, 5\}$. Coefficients minimize the method's empirical squared

loss, $\sum_r (y_r - b_0^\top \phi(x_r) - p_r\, b_1^\top \phi(x_r))^2$, solved exactly by least squares (minimum-norm solution when the design is rank-deficient, as it is for DCML). DCMOML minimizes the loss averaged over both query indices in A_i. Using an exactly solvable learner removes optimization error from the comparison, so differences between methods reflect information design and supervision.

Other baselines. EB-GLS models $\Theta_i \sim \mathcal{N}(\mu, \Sigma)$ independently of prices, with a Gaussian likelihood whose standard deviation is the known noise level 0.5; (μ, Σ) are estimated by EM on the training tasks, and the estimate is the posterior mean. SHARED is pooled OLS. TASK-OLS is OLS on each task's own K observations. With $K = 2$, its slope error is $(\epsilon_{ia} - \epsilon_{ib})/(p_{ia} - p_{ib})$, and the price gap has positive density at zero, so its mean squared error is infinite. Sample averages are therefore unstable, and we report the median squared error.

Oracle. The oracle is the Bayes floor for DCMOML's information set, $\mathbb{E}[\Theta_i \mid X_i^{-A_i}]$, computed by numerical integration under the known DGP. The posterior is evaluated on a 201×201 grid over (θ^0, θ^1) (six prior standard deviations). The price likelihood integrates out the manager's target on a 1,200-point grid, and the likelihood of the unmasked demands is included. Refining the grids changes the oracle MSEs by less than 10^{-4}. The oracle uses no learner, so it independently checks how close DCMOML comes to its identified target.

Protocol. For each $K \in \{2, 3, 4, 5\}$ and $\sigma_c \in \{0, 0.1, 0.2\}$, and for each of 10 seeds, we draw 20,000 training tasks and 10,000 new test tasks, fit every method on the training tasks, and compute MSE for θ_i^0 and θ_i^1 on the test tasks. We report the mean across seeds ± 1.96 standard errors. Tables 2 and 3 give all results. Table 4 varies the number of training tasks under high confounding for $K = 2$ and $K = 5$, comparing DCMOML and META. Code to reproduce every number is in the supplementary material.

Table 2: Synthetic study: intercept MSE (θ_i^0). Mean over 10 seeds ± 1.96 SE; each seed trains on 20,000 tasks and evaluates on 10,000 new tasks. †Median squared error, because per-task OLS has no finite mean squared error (see text).

K	Method	$\sigma_c = 0$ (HC)	$\sigma_c = 0.1$ (MC)	$\sigma_c = 0.2$ (LC)
2	DCMOML	0.647 ± 0.015	0.775 ± 0.028	0.953 ± 0.055
	META	2.004 ± 0.039	0.779 ± 0.034	0.428 ± 0.011
	DCML	96.4 ± 0.4	96.3 ± 0.4	94.4 ± 0.6
	DCUOML	8.846 ± 0.440	13.5 ± 0.8	15.8 ± 0.9
	EB-GLS	10.1 ± 0.2	5.514 ± 0.112	2.351 ± 0.045
	SHARED	25.8 ± 0.2	15.1 ± 0.1	5.728 ± 0.102
	TASK-OLS†	24.5	24.9	25.1
	Oracle	0.599 ± 0.005	0.716 ± 0.006	0.849 ± 0.005
3	DCMOML	0.392 ± 0.011	0.385 ± 0.014	0.384 ± 0.010
	META	0.769 ± 0.047	0.374 ± 0.018	0.319 ± 0.010
	DCML	96.7 ± 0.5	96.2 ± 0.4	94.3 ± 0.7
	DCUOML	3.357 ± 0.191	4.664 ± 0.205	5.271 ± 0.188
	EB-GLS	3.685 ± 0.076	2.403 ± 0.047	1.293 ± 0.019
	SHARED	25.8 ± 0.2	15.1 ± 0.1	5.635 ± 0.042
	TASK-OLS†	8.33	8.44	8.29
	Oracle	0.334 ± 0.002	0.336 ± 0.003	0.339 ± 0.003
4	DCMOML	0.315 ± 0.004	0.319 ± 0.005	0.323 ± 0.008
	META	0.462 ± 0.023	0.300 ± 0.013	0.288 ± 0.007
	DCML	96.6 ± 0.3	96.5 ± 0.7	94.5 ± 0.6
	DCUOML	1.854 ± 0.102	2.539 ± 0.120	2.904 ± 0.110
	EB-GLS	1.665 ± 0.034	1.298 ± 0.032	0.855 ± 0.017
	SHARED	25.8 ± 0.1	15.1 ± 0.1	5.715 ± 0.089
	TASK-OLS†	4.88	4.89	4.9
	Oracle	0.260 ± 0.002	0.270 ± 0.002	0.275 ± 0.002
5	DCMOML	0.261 ± 0.005	0.281 ± 0.007	0.286 ± 0.004
	META	0.334 ± 0.025	0.267 ± 0.007	0.264 ± 0.004
	DCML	96.7 ± 0.4	96.1 ± 0.3	93.7 ± 0.7
	DCUOML	1.447 ± 0.074	1.616 ± 0.079	1.851 ± 0.242
	EB-GLS	0.980 ± 0.016	0.833 ± 0.020	0.637 ± 0.014
	SHARED	25.8 ± 0.2	15.0 ± 0.2	5.750 ± 0.061
	TASK-OLS†	3.48	3.43	3.5
	Oracle	0.221 ± 0.002	0.241 ± 0.002	0.250 ± 0.002

Table 3: Synthetic study: slope MSE (θ_i^1). Mean over 10 seeds $\pm$ 1.96 SE; each seed trains on 20,000 tasks and evaluates on 10,000 new tasks. †Median squared error, because per-task OLS has no finite mean squared error (see text).

K	Method	$\sigma_c = 0$ (HC)	$\sigma_c = 0.1$ (MC)	$\sigma_c = 0.2$ (LC)
2	DCMOML	0.0077 ± 0.0005	0.0089 ± 0.0010	0.0107 ± 0.0010
	META	0.0783 ± 0.0020	0.0267 ± 0.0015	0.0113 ± 0.0003
	DCML	3.962 ± 0.017	4.066 ± 0.016	4.389 ± 0.019
	DCUOML	0.349 ± 0.019	0.531 ± 0.034	0.614 ± 0.025
	EB-GLS	0.408 ± 0.007	0.223 ± 0.004	0.0936 ± 0.0016
	SHARED	1.001 ± 0.008	0.574 ± 0.006	0.203 ± 0.004
	TASK-OLS†	0.98	1	1.05
	Oracle	0.0059 ± 0.0000	0.0071 ± 0.0001	0.0084 ± 0.0001
3	DCMOML	0.0073 ± 0.0004	0.0084 ± 0.0004	0.0092 ± 0.0001
	META	0.0275 ± 0.0018	0.0111 ± 0.0006	0.0090 ± 0.0003
	DCML	3.975 ± 0.019	4.081 ± 0.014	4.391 ± 0.024
	DCUOML	0.128 ± 0.008	0.181 ± 0.007	0.200 ± 0.007
	EB-GLS	0.152 ± 0.003	0.0991 ± 0.0020	0.0522 ± 0.0007
	SHARED	1.002 ± 0.006	0.577 ± 0.005	0.201 ± 0.003
	TASK-OLS†	0.33	0.334	0.339
	Oracle	0.0053 ± 0.0000	0.0069 ± 0.0001	0.0078 ± 0.0001
4	DCMOML	0.0066 ± 0.0001	0.0085 ± 0.0001	0.0091 ± 0.0003
	META	0.0153 ± 0.0011	0.0091 ± 0.0006	0.0091 ± 0.0004
	DCML	3.982 ± 0.013	4.082 ± 0.023	4.413 ± 0.019
	DCUOML	0.0690 ± 0.0036	0.0958 ± 0.0045	0.106 ± 0.004
	EB-GLS	0.0700 ± 0.0015	0.0541 ± 0.0012	0.0351 ± 0.0007
	SHARED	1.001 ± 0.005	0.576 ± 0.006	0.204 ± 0.003
	TASK-OLS†	0.194	0.195	0.2
	Oracle	0.0048 ± 0.0000	0.0068 ± 0.0001	0.0077 ± 0.0000
5	DCMOML	0.0056 ± 0.0001	0.0080 ± 0.0002	0.0086 ± 0.0002
	META	0.0102 ± 0.0010	0.0083 ± 0.0002	0.0085 ± 0.0002
	DCML	3.970 ± 0.017	4.070 ± 0.012	4.399 ± 0.023
	DCUOML	0.0517 ± 0.0031	0.0606 ± 0.0028	0.0672 ± 0.0078
	EB-GLS	0.0414 ± 0.0008	0.0351 ± 0.0008	0.0260 ± 0.0006
	SHARED	1.003 ± 0.007	0.576 ± 0.006	0.205 ± 0.002
	TASK-OLS†	0.137	0.136	0.142
	Oracle	0.0043 ± 0.0000	0.0068 ± 0.0001	0.0076 ± 0.0001

Table 4: Synthetic study, high confounding: MSE as the number of training tasks grows (mean over 5 seeds for $K = 2$ and 3 seeds for $K = 5$). The oracle does not depend on the number of training tasks.

		Intercept MSE			Slope MSE		
K	Tasks	DCMOML	META	Oracle	DCMOML	META	Oracle
2	2,000	0.948	2.107	0.597	0.0203	0.0837	0.0059
	5,000	0.917	2.163	0.596	0.0155	0.0863	0.0060
	20,000	0.637	2.020	0.601	0.0074	0.0789	0.0059
	100,000	0.607	2.014	0.599	0.0063	0.0795	0.0059
	400,000	0.601	1.993	0.600	0.0061	0.0787	0.0060
5	5,000	0.397	0.403	0.218	0.0098	0.0131	0.0043
	20,000	0.259	0.353	0.223	0.0056	0.0110	0.0044
	50,000	0.240	0.297	0.220	0.0049	0.0089	0.0043

D Implementation Details for Retail Experiments

This appendix documents the data processing, objective, architecture and optimization used in Section 6.2. The released code reproduces every number in Section 6.2 and in this appendix.

D.1 Data processing

The processed UCI Online Retail panel has one row per product, transaction price and day with at least one sale, giving the number of units sold. Days without sales are not recorded, so a product's "exposure" at a price counts sale days, and average daily demand is an average over sale days. Each product's 1024-dimensional title embedding, computed from the product description with the sentence-transformers model `all-roberta-large-v1` [Reimers and Gurevych, 2019, Sentence-Transformers, 2021], is attached as Z_i. The UCI Online Retail dataset is distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license [Chen, 2015], and the embedding model under the Apache 2.0 license.

Static-Top3. For each product and price we count the number of sale days at that price and compute average daily demand over those days. Prices are ranked by this count within product; products with at least three distinct prices are kept ($n = 2{,}833$). Ranks 1–2 form the training pairs and rank 3 is held out. Among these products, the top two prices account on average for 89.7% of the sale days at the top three prices.

Exposure-Sequence. Rows with non-positive prices are removed. Daily sales are summed per product and day, and the day's price is the median transaction price. Consecutive sale days with the same price form a run, giving (p_{ik}, e_{ik}, D_{ik}) = (run price, number of sale days, average daily sales). Products with at least three runs are kept ($n = 3{,}132$); the first two runs are used for training and the third is held out.

D.2 Exposure-weighted supervision

Each outcome D_{ik} averages over e_{ik} days, so longer exposures are less noisy. Let $\mathcal{Q}_i = \{1, 2\}$ be the supervised query indices of product i. With per-product normalized weights $\tilde{w}_{ik} = e_{ik} / \sum_{j \in \mathcal{Q}_i} e_{ij}$, the per-product loss is

$$\mathcal{L}_i \;=\; \sum_{k \in \mathcal{Q}_i} \tilde{w}_{ik} \left(\hat{D}_{ik} - D_{ik}\right)^2, \qquad \hat{D}_{ik} = \hat{\theta}_i^0 + \hat{\theta}_i^1 p_{ik},$$

where $(\hat{\theta}_i^0, \hat{\theta}_i^1)$ comes from the method's own information set (for META, from the support point opposite to query k). The training objective is the minibatch average of $\mathcal{L}_i$, and the same loss on validation products is used for early stopping. At evaluation, META's two support-specific estimates are averaged before predicting observation 3.

D.3 Architecture and optimization

All methods predict (θ_i^0, θ_i^1) with the same MLP,

$$\texttt{Linear}(d_{\text{in}}, 256) \to \texttt{ReLU} \to \texttt{Linear}(256, 256) \to \texttt{ReLU} \to \texttt{Linear}(256, 2),$$

where d_{in} is the method-specific input dimension (1024 for SHARED; 1026 for DCMOML and META on Static-Top3; 1028 for DCMOML and 1027 for META on Exposure-Sequence). Inputs are used without normalization. We use AdamW with weight decay 10^{-5}, batch size 512, learning rate 10^{-3} (Static-Top3) or 3×10^{-4} (Exposure-Sequence), at most 100 epochs, and early stopping with patience 10 on the validation loss, restoring the best checkpoint. Seeds $1, \ldots, 100$ set the product split and the initialization; all methods share the split within a seed. Confidence intervals are mean $\pm\, 1.96\times$ (standard deviation across seeds)$/\sqrt{100}$.

Training instability. On Static-Top3, META's fitted slopes are mostly positive in 10 of 100 seeds (never for DCMOML or SHARED). Section 6.2 reports the median paired difference, which is robust to these seeds. On Exposure-Sequence all three methods have seeds with near-zero average predicted demand (META 17, SHARED 13, DCMOML 3): seeds in which the exposure-weighted average predicted demand at the held-out price is below 10% of the exposure-weighted average training demand.

PER-TASK. Weighted least squares per product on all non-holdout prices (weights e_{ik}), solved in closed form; it has no randomness.

E The DCMOML risk as design-weighted regression of two-point estimates

Suppress the task index and use the notation of Appendix B: $X = X^{-A}$, $A = \{a, b\}$ and $Q(X) = \frac{1}{2}P_aP_a^\top + \frac{1}{2}P_bP_b^\top$. Let H be the 2×2 matrix with rows $P_a^\top$ and $P_b^\top$, which is invertible because $p_a \neq p_b$, and let

$$\widetilde{\Theta} = H^{-1}\begin{pmatrix} D_a \\ D_b \end{pmatrix}$$

be the two-point least-squares estimate of Θ, the coefficients of the line through the two selected observations. Both H and Q are functions of X.

Proposition 1. *For every measurable g, pointwise,*

$$\frac{1}{2}\sum_{k\in A}\big(D_k - P_k^\top g(X)\big)^2 = \big(\widetilde{\Theta} - g(X)\big)^\top Q(X)\big(\widetilde{\Theta} - g(X)\big). \tag{17}$$

Consequently, the DCMOML population risk is $\mathcal{L}(g) = \mathbb{E}\big[(\widetilde{\Theta} - g(X))^\top Q(X)(\widetilde{\Theta} - g(X))\big]$. Moreover, under the assumptions of Theorem 1,

$$H^{-1}\begin{pmatrix} \mathbb{E}[D_a \mid X] \\ \mathbb{E}[D_b \mid X] \end{pmatrix} = g^*(X).$$

If $\widetilde{\Theta}$ is integrable, this also gives $\mathbb{E}[\widetilde{\Theta} \mid X] = g^*(X)$. The weighted-loss identity does not require this additional integrability.

Proof. By construction $P_k^\top\widetilde{\Theta} = D_k$ for $k \in A$, so $D_k - P_k^\top g(X) = P_k^\top(\widetilde{\Theta} - g(X))$, and

$$\frac{1}{2}\sum_{k\in A}\big(P_k^\top(\widetilde{\Theta} - g)\big)^2 = (\widetilde{\Theta} - g)^\top\Big(\tfrac{1}{2}\sum_{k\in A}P_kP_k^\top\Big)(\widetilde{\Theta} - g) = (\widetilde{\Theta} - g)^\top Q(\widetilde{\Theta} - g).$$

Since κ is uniform on A given X and independent of the outcomes given X, $\mathcal{L}(g) = \mathbb{E}\big[\frac{1}{2}\sum_{k\in A}(D_k - P_k^\top g(X))^2\big]$, which gives the expression for $\mathcal{L}$. Finally, $\mathbb{E}[D_k \mid X] = P_k^\top g^*(X)$ for $k \in A$ by (13) and the demand model. Multiplying these two conditional means by H^{-1} gives $H^{-1}H\,g^*(X) = g^*(X)$. If $\widetilde{\Theta}$ is integrable, taking its conditional expectation yields the same expression because H is X-measurable. □

DCMOML is therefore a regression of each task's two-point estimate $\widetilde{\Theta}$ on its masked input X, weighted by the design matrix $Q(X)$: the fitted map shrinks the noisy task-level estimate toward what the cross-task information in X predicts. Writing $\Delta = \widetilde{\Theta} - g(X)$, $\bar{p} = (p_a + p_b)/2$ and $h = (p_b - p_a)/2$, the weighted loss is

$$\Delta^\top Q\Delta = (\Delta_0 + \bar{p}\,\Delta_1)^2 + h^2\Delta_1^2,$$

the squared error at the midpoint price plus the slope error scaled by the half-gap. The slope of $\widetilde{\Theta}$ is least reliable when the two prices are close, and this is exactly where it receives the least weight. The weighting is essential. Given X, the slope of $\widetilde{\Theta}$ has error $(\epsilon_b - \epsilon_a)/(p_b - p_a)$; when the price gap has positive density at zero, as with Gaussian prices, $\widetilde{\Theta}$ has infinite variance and need not even be integrable, so an unweighted regression of $\widetilde{\Theta}$ on X has infinite risk. By (17), the weighted risk is finite for every g satisfying the integrability requirement of Theorem 1.

NeurIPS Paper Checklist

1. **Claims.** Answer: Yes. Justification: The main claims are stated in the introduction as three explicit contributions: introducing DCMOML with identification and consistency guarantees, establishing maximal uniform validity for its two-point squared-loss objective, and evaluating the method on synthetic and real-world data. The scope of the causal claims is tied to the model assumptions in Section 3 and the information-design conditions in Section 5.
2. **Limitations.** Answer: Yes. Justification: Section 7 discusses the scope of the guarantees, the limitations of the assumptions and experiments, and directions for future work.
3. **Theory assumptions and proofs.** Answer: Yes. Justification: The formal assumptions are stated in Section 3: the linear demand model, the task heterogeneity decomposition, cross-task price endogeneity, mean-zero, finite-variance shocks independent of each other and of (Z_i, Θ_i), the local exogeneity Assumption 1 with two distinct selected prices, and the moment condition, Assumption 2. Theorem 1 gives identification and consistency conditions (realizability and uniform convergence are stated in the theorem), Theorem 2 gives maximal validity for the squared-loss objective, and Appendix B provides complete proofs together with remarks on weaker conditions.
4. **Experimental result reproducibility.** Answer: Yes. Justification: Appendix C specifies the synthetic data-generating process, learner, baselines, oracle computation, sample sizes and seeds. Appendix D specifies the retail preprocessing, objective, architecture, optimization and seeds. The released code reproduces every number in Section 6.
5. **Open access to data and code.** Answer: Yes. Justification: The real-world dataset is the publicly available UCI Online Retail dataset. Code for the synthetic and retail experiments, including the processed retail panel, is provided as supplementary material.
6. **Experimental setting/details.** Answer: Yes. Justification: Section 6 describes the tasks, baselines, protocols and metrics; Appendices C and D give the full settings.
7. **Experiment statistical significance.** Answer: Yes. Justification: Synthetic results are means over 10 seeds with 95% intervals (Figure 1, Tables 2–3). Retail results are means over 100 seeds with 95% intervals, together with paired seed-level differences and the share of seeds in which each comparison favors DCMOML (Table 1).
8. **Compute resources.** Answer: Yes. Justification: All experiments run on CPUs. The full synthetic study takes about 80 minutes on two cores; one retail seed (all methods) takes about one minute on one core. No GPUs or clusters were used.
9. **Code of ethics.** Answer: Yes. Justification: The work studies causal demand estimation for pricing and uses a public retail transaction dataset to illustrate the methodology. The paper does not involve deception, private personal data collection, or interventions on users. The contribution adheres to the NeurIPS Code of Ethics.
10. **Broader impacts.** Answer: Yes. Justification: This paper develops methods for causal demand estimation under endogenous pricing, with the goal of improving the reliability of price-response estimates in settings where observational pricing data are confounded by latent demand factors. More accurate causal demand learning can benefit firms and consumers by reducing reliance on biased observational correlations, supporting better inventory and pricing decisions, and enabling more transparent evaluation of how prices affect demand across heterogeneous contexts.

    At the same time, pricing methods have important societal implications. Improved demand estimation could be used to implement more effective price discrimination, including pricing strategies that exploit consumer heterogeneity or disproportionately affect vulnerable groups. Although our method is intended as an estimation tool rather than a complete pricing policy, downstream deployment in real pricing systems should be accompanied by fairness audits, monitoring for distribution shift, human oversight, and explicit constraints against discriminatory or exploitative pricing practices. Particular care is needed when context covariates may encode protected or sensitive attributes, either directly or through proxies such as geography or demographics.

    Our experiments use synthetic data and a public retail transaction dataset, and we do not collect new human-subject data or deploy the method in a live pricing environment. The proposed approach is therefore not expected to create direct immediate harms. The main risks arise from future applications in automated pricing systems. We encourage practitioners to evaluate not only

predictive or revenue performance, but also consumer welfare, fairness across groups, robustness to changing market conditions, and compliance with applicable laws and platform policies.

11. **Safeguards.** Answer: NA. Justification: The paper does not release models or datasets with a high risk of misuse requiring controlled access. Potential harms from downstream pricing applications and possible mitigation measures are discussed in the broader impact statement above.
12. **Licenses for existing assets.** Answer: Yes. Justification: The UCI Online Retail dataset (CC BY 4.0) and the sentence-transformers model `all-roberta-large-v1` used for the product-title embeddings (Apache 2.0) are cited, and their licenses are stated in Appendix D.1.
13. **New assets.** Answer: Yes. Justification: The paper provides experimental code and processed task definitions, described in Appendices C and D and in the accompanying READMEs. The code is released under the MIT License, and the processed data derived from the UCI Online Retail dataset under CC BY 4.0, as stated in the release.
14. **Crowdsourcing and research with human subjects.** Answer: NA. Justification: The paper does not involve crowdsourcing, annotation by human workers, user studies, surveys, or newly collected human-subject data. The empirical evaluation uses synthetic data and an existing public retail transaction dataset.
15. **Institutional review board approvals.** Answer: NA. Justification: The paper does not involve human-subject experiments, interventions, surveys, or collection of identifiable private information. Therefore IRB approval does not appear applicable.
16. **LLM usage.** Answer: Yes. Justification: We used large language models (LLMs) as assistants. ChatGPT 5.2–5.5 and Gemini 3.0 helped draft prose, edit for clarity, generate LaTeX scaffolding, and implement portions of the experimental pipeline. In preparing this revision, GPT-6 and Claude Opus 5.5 were used to check the proofs and suggest expositional improvements. Claude Opus 5.5 was also used as a coding assistant to implement and run the synthetic experiments, the scoring of the retail experiments, and the code that produces the figures and reported numbers. The methodology and all decisions about the experiments and claims were made by the authors. All code, proofs, analyses, and interpretations were checked by the authors, who take full responsibility for the scientific integrity and contents of the manuscript.