Transfer Learning for Contextual Joint Assortment–Pricing: Multi-Source Utility Shift under Multinomial Logit Model

Elynn Chen[♯], Xi Chen[†], Yi Zhang[♭]

[♯] [†] Leonard N. Stern School of Business, New York University
[♭] Fu Foundation School of Engineering, Columbia University

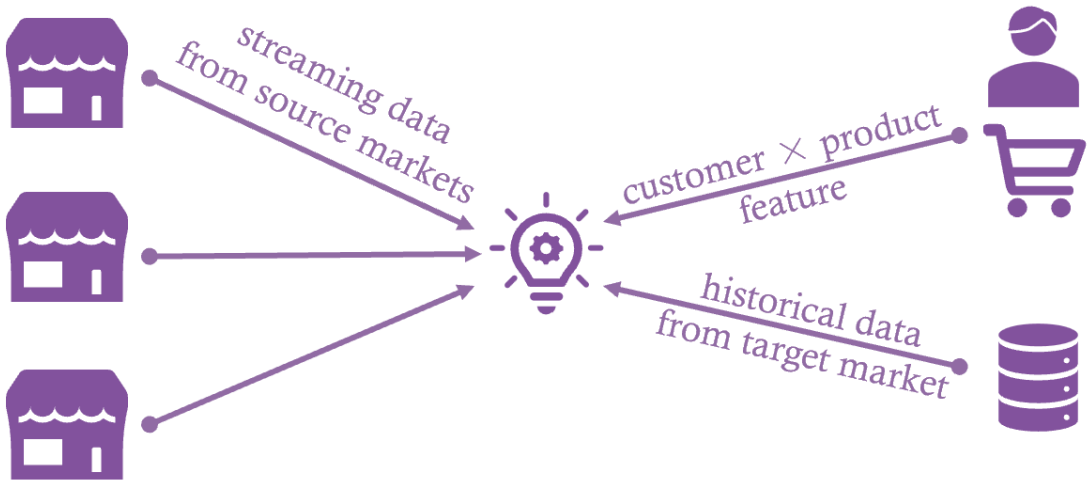
October 24, 2025

Motivation

Background. Accurately learning consumer demand is central to dynamic decision making in digital marketplaces. In many applications, sellers must determine *both* which products to display and what prices to post to sequentially arriving customers.

Gap. Prior work typically treats assortment or pricing alone. Single-market algorithms learn from scratch, wasting useful information from auxiliary markets.

Our goal. Bring *transfer* to contextual *joint* assortment–pricing under MNL bandit feedback: leverage multiple sources to accelerate learning in a designated target market.



Related Work: Assortment and Pricing under MNL

Online assortment *or* pricing under MNL. Non-contextual MNL bandits (Agrawal et al. 2017, Cheung & Simchi-Levi 2017); contextual MNL bandit (Agrawal et al. 2018, Oh & Iyengar 2019, Chen et al. 2020, Oh & Iyengar 2021).

Online assortment *and* pricing under MNL. Miao & Chao (2021) analyzes a non-contextual formulation and design a cycle-based TS policy whose regret scales with the catalog size N; we treat the contextual case, and our regret depends only on the assortment size K. Erginbas et al. (2025) considers contextual formulation; our approach focuses on cross-market transfer, addressing the attendant estimation issues in aggregating and debiasing. Even without transfer, our obtained regret bound scales as $\widetilde{O}(\sqrt{d})$, strictly sharper than their $\widetilde{O}(d)$ dependence.

Related Work: Bandits for discrete and combinatorial decisions.

Contextual bandits and UCB algorithm. Contextual linear and generalized linear bandits (Abbasi-Yadkori et al. 2011, Dani et al. 2008, Rusmevichientong & Tsitsiklis 2010, Filippi et al. 2010, Li et al. 2017). The Upper Confidence Bound (UCB) method and its variants (Chen et al. 2013, Kveton et al. 2015). UCB algorithms in MNL Bandits (Agrawal et al. 2017, Chen et al. 2020, Oh & Iyengar 2021).

Our distinction. Existing literature is less developed in the regime where assortment selection must be jointly coupled with continuous pricing under bandit feedback. We work within the established MNL paradigm, adopt a UCB-type approach, and, crucially, address the joint, contextual decision space that inherently mixes discrete assortments with continuous prices.

Related Work: Transfer and meta learning

Transfer and meta learning. Prior studies clarify when and how auxiliary data improve estimation or linear bandits: supervised transfer with sparse-contrast (Bastani 2021); meta-dynamic pricing with a shared prior and observed linear demand (Bastani et al. 2022); covariate-shift transfer from one offline source with invariant rewards (Cai et al. 2024); robust multi-task bandits that trim outliers but omit price optimization (Xu & Bastani 2024); and estimation-only transfer (Li et al. 2022, Tian & Feng 2022, Liu et al. 2023).

Our distinction. They leave open the design of transfer mechanisms tailored to MNL bandits under mixed decision spaces and multi-source utility shifts; nor do they incorporate revenue-maximising price choice.

Contribution

Model and algorithmic innovation. We propose an *aggregate–then–debias* pipeline tailored to the utility-shift structure. Building on the estimates, we design a *two-radius UCB* decision rule. The first radius tracks *self-normalized variance*; the second radius upper-bounds *transfer bias* inherited from cross-market aggregation.

Theoretical innovation. We derive matching upper and lower bounds with a decomposition that isolates two statistical sources of uncertainty introduced by transfer: a variance term governed by self-normalized information, and a transfer-bias term driven by cross-market heterogeneity.

Practical guidance. We obtain managerial insights on *when* and *how much* transfer helps. Gains accrue with H as long as cross-market heterogeneity s_0 remains sufficiently *sparse*; once discrepancies become diffuse, the bias term dominates and transfer benefits taper off.

Model, Choice and Revenue

Assortments. Catalog set $[N]$, feasible sets $\mathcal{S}_K := \{S \subseteq [N] : |S| \le K\}$.
Decision. Choose $(S_t, \boldsymbol{p}_t)$ with $S_t \in \mathcal{S}_K$, $\boldsymbol{p}_t \in \mathbb{R}^N$.
Utility model. $v_{it} = \langle \boldsymbol{x}_{it}, \boldsymbol{\theta} \rangle - \langle \boldsymbol{x}_{it}, \boldsymbol{\gamma} \rangle \, p_{it} + \varepsilon_{it}, \quad i \in S_t,.$
MNL choice probabilities.

$$q_t(i \mid S_t, \boldsymbol{p}_t) = \frac{\exp(v_{it})}{1 + \sum_{\ell \in S_t} \exp(v_{\ell t})}, \qquad i \in S_t. \tag{1}$$

Expected revenue.

$$R_t(S_t, \boldsymbol{p}_t) := \sum_{i \in S_t} p_{it} \, q_t(i \mid S_t, \boldsymbol{p}_t). \tag{2}$$

Clairvoyant Policy.

$$(S_t^*, \boldsymbol{p}_t^*) \in \operatorname*{argmax}_{S \in \mathcal{S}_K, \ \boldsymbol{p} \in \mathbb{R}^N} \sum_{i \in S} p_{it} \, \frac{\exp(v_{it})}{1 + \sum_{j \in S} \exp(v_{jt})}. \tag{3}$$

Regret over horizon T.

$$\mathrm{Regret}(T; \pi) = \sum_{t=1}^{T} R_t(S_t^*, \boldsymbol{p}_t^*) - \sum_{t=1}^{T} R_t(S_t, \boldsymbol{p}_t).$$

Cross-market Transfer

Target $(^{(0)})$. $v_{it}^{(0)} = \langle x_{it}^{(0)}, \theta^{(0)} \rangle - \langle x_{it}^{(0)}, \gamma^{(0)} \rangle p_{it} + \varepsilon_{it}$.

Source $(^{(h)})$. $v_{it}^{(h)} = \langle x_{it}^{(h)}, \theta^{(h)} \rangle - \langle x_{it}^{(h)}, \gamma^{(h)} \rangle p_{it} + \varepsilon_{it}, \quad h \in [H]$

Assumption (Homogeneous Covariates with Bounded Spectrum)

For $h \in \{0\} \cup [H]$, $x_{it}^{(h)} \overset{\text{i.i.d.}}{\sim} \mathcal{P}_x$ supported on bounded $\mathcal{X} \subset \mathbb{R}^d$, $\mathbb{E}[x_{it}] = 0$, $\Sigma = \mathbb{E}[x_{it} x_{it}^\top]$ with $0 < C_{\min} \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq C_{\max} < \infty$.

Assumption (Task Similarity)

The maximum l_0-norm of the difference between target and source coefficients is bounded:

$$\max_{h \in [H]} \left(\|\nu^{(0)} - \nu^{(h)}\|_0 \right) \leq s_0.$$

Optimisitic Utility Cobstruction

Construct $\mathcal{C}_m$, an ellipsoidal confidence region containing $\boldsymbol{\nu}^{(0)}$ w.h.p.

$$\bar{v}_{it}^{(0)}(p) \;=\; \langle \boldsymbol{x}_{it}^{(0)}, \widehat{\boldsymbol{\theta}} \rangle \;-\; \langle \boldsymbol{x}_{it}^{(0)}, \widehat{\boldsymbol{\gamma}} \rangle \, p \;+\; u_{it}(p),$$

with a *two-radius* bonus

$$u_{it}(p) = \alpha_m \, \|\widetilde{\boldsymbol{x}}_{it}^{(0)}(p)\|_{W_{m-\mathbf{1}}^{-\mathbf{1}}} \;+\; \beta_m \, \|\widetilde{\boldsymbol{x}}_{it}^{(0)}(p)\|_{\infty}. \tag{4}$$

$\bar{v}_{it}^{(0)}(p)$ may conflicts with positive price sensitivity Assumption. Enforce decreasing L_0-Lipschitz via:

Lemma (Monotone–Lipschitz envelope)

If $\bar{v}_{it}(p) \geq v_{it}(p)$ on $[0, \bar{P}]$, then $\widetilde{v}_{it}(p) := \min_{p' \leq p} \left\{ \bar{v}_{it}^{(0)}(p') - L_0 \, (p - p') \right\}$ is decreasing, L_0-Lipschitz, and $v_{it}(p) \leq \widetilde{v}_{it}(p)$ for all p.

Choose

$$(S_t, \boldsymbol{p}_t) \;\in\; \underset{S \in \mathcal{S}_K, \; \boldsymbol{p} \in \mathbb{R}^K}{\operatorname{argmax}} \; \frac{\sum_{i \in S} p_i \exp(\widetilde{v}_{it}(p_i))}{1 + \sum_{j \in S} \exp(\widetilde{v}_{jt}(p_j))},$$

which reduces to a one-dimensional fixed point problem Wang (2012).

Per-period & Rolling Information Matrix

Per-period Fisher Information Matrix:

$$\mathcal{I}_t^{(h)}(\nu) = \sum_{i \in S_t^{(h)}} q_{it}^{(h)}(\nu)\, \widetilde{x}_{it}^{(h)} \widetilde{x}_{it}^{(h)\top} - \sum_{i \in S_t^{(h)}} \sum_{j \in S_t^{(h)}} q_{it}^{(h)}(\nu)\, q_{jt}^{(h)}(\nu)\, \widetilde{x}_{it}^{(h)} \widetilde{x}_{jt}^{(h)\top}.$$

Rolling Fisher within episode m:

$$V_t^{(h)} = \sum_{u=\tau_{m-1}+1}^{t} \mathcal{I}_u^{(h)}(\widehat{\nu}_m), \quad \forall h \in 0 \cup [H].$$

$V_t^{(0)}$ is used to check an identifiability gate that triggers forced exploration.

Algorithm 1: Subroutine: OfferAssortmentAndPrice

Input: $t,\ \tau_m,\ q_m,\ V_t^{(\mathbf{0})},\ \widetilde{C}_{\min},\ K,\ \bar{P},\ \widehat{\boldsymbol{\nu}}_m,\ \alpha_m,\ \beta_m,\ W_{m-\mathbf{1}}$

Output: Decision $(S_t^{(\mathbf{0})}, \boldsymbol{p}_t^{(\mathbf{0})})$

if $\tau_m - t \le q_m$ ***and*** $\lambda_{\min}(V_t^{(\mathbf{0})}) \le \dfrac{K q_m \widetilde{C}_{\min}}{2}$ **then**

  Randomly choose $S_t^{(\mathbf{0})} \in \mathcal{S}_K$ and $\boldsymbol{p}_t^{(\mathbf{0})} \sim \mathrm{Uniform}([0, \bar{P}]^N)$;

else

  $(S_t^{(\mathbf{0})}, \boldsymbol{p}_t^{(\mathbf{0})}) \leftarrow \underset{S \in \mathcal{S}_K,\ \boldsymbol{p} \in \mathbb{R}^N}{\mathrm{argmax}} \widetilde{R}_t(S, \boldsymbol{p}; \alpha_m, \beta_m, W_{m-\mathbf{1}})$;

return $(S_t^{(\mathbf{0})}, \boldsymbol{p}_t^{(\mathbf{0})})$;

Episodic Information Matrix

Episodic Fisher Information Matrix:

$$W_{m-1} := V^{(0)}_{\tau_{m-1}} + \sum_{h=1}^{H} \omega_h V^{(h)}_{\tau_{m-1}}. \tag{5}$$

W_{m-1} appears only in the variance bonus $\|\widetilde{x}\|_{W^{-1}_{m-1}}$ of (4).

(i) Homogeneous covariates (Assumption 2.1). Take $\omega_h = 1$. Then $\lambda_{\min}(W_{m-1})$ grows $\widetilde{\Omega}(1+H)$, and $\|\widetilde{x}\|_{W^{-1}_{m-1}}$ contracts at $\widetilde{\mathcal{O}}(1/\sqrt{1+H})$.

(ii) Heterogeneous covariates. Temper $\omega_h V^{(h)}_{\tau_{m-1}}$ by: *sample reweighting* (density-ratio correction inside each $V^{(h)}$); or *market weights* $\omega_h \in [0,1]$ from mismatch scores (e.g., $\omega_h = \{1 + \widehat{\chi}^2(P_0 \| P_h)\}^{-1}$). This preserves PD while guarding bias.

Aggregate-then-debias Pipeline

The horizon T is partitioned into episodes $m = 1, 2, \ldots, M$, where episode m has length $\tau_m = 2^{m-1}$. Thus $M = \lceil \log_2 T \rceil$ and parameter updates occur only $\mathcal{O}(\log T)$ times.
At the start of each episode m, we update the parameter estimates using data from the *preceding* episode:

(i) **Aggregate with source market data.**

(ii) **Debias with target market data.**

Algorithm 2: Subroutine: AggregateThenDebias

Input: Source losses $\{\mathcal{L}_t^{(h)}\}_{h\in[H],\, t\in\mathcal{T}_{m-1}}$, target losses $\{\mathcal{L}_t^{(0)}\}_{t\in\mathcal{T}_{m-1}^{(0)}}$, regularization λ_m

Output: Episode-m parameter $\widehat{\nu}_m$

// (i) weighted aggregate on sources from episode $m-1$

$\widehat{\nu}_m^{(ag)} \leftarrow \operatorname{argmin}_{\nu\in\mathbb{R}^{2d}} \frac{1}{H|\mathcal{T}_{m-1}|} \sum_{h\in[H]} \sum_{t\in\mathcal{T}_{m-1}} \mathcal{L}_t^{(h)}(\nu)$;

// (ii) debias with target data from episode $m-1$

$\widehat{\delta}_m \leftarrow \operatorname{argmin}_{\delta\in\mathbb{R}^{2d}} \left(\frac{1}{|\mathcal{T}_{m-1}^{(0)}|} \sum_{t\in\mathcal{T}_{m-1}^{(0)}} \mathcal{L}_t^{(0)}(\widehat{\nu}_m^{(ag)} + \delta) + \lambda_m \|\delta\|_1 \right)$;

return $\widehat{\nu}_m \leftarrow \widehat{\nu}_m^{(ag)} + \widehat{\delta}_m$;

Complete Algorithm

TJAP-CWF: Transfer joint assortment–pricing with cross-market weighted information

Algorithm 3: TJAP-CWF

Input: Streaming data $\{\{\boldsymbol{x}_{it}^{(h)}\}_{i\in[N]}, S_t^{(h)}, \boldsymbol{p}_t^{(h)}, \boldsymbol{y}_t^{(h)}\}_{t\geq \mathbf{1}}$ for $h \in \{0\} \cup [H]$

Initialize: $V_{\mathbf{0}}^{(\mathbf{0})} \leftarrow 0_{\mathbf{2}d\times\mathbf{2}d}$; $V_{\mathbf{0}}^{(h)} \leftarrow 0_{\mathbf{2}d\times\mathbf{2}d}$ for all $h \in [H]$

for $t \in [2d]$ **do**
- Randomly choose $S_t^{(\mathbf{0})} \in \mathcal{S}_K$ and $\boldsymbol{p}_t^{(\mathbf{0})} \sim \text{Uniform}([0, \bar{P}]^N)$;
- $V_t^{(\mathbf{0})} \leftarrow V_{t-\mathbf{1}}^{(\mathbf{0})} + \frac{\mathbf{1}}{K^{\mathbf{2}}} \sum_{i\in S_t^{(\mathbf{0})}} \widetilde{\boldsymbol{x}}_{it}^{(\mathbf{0})} (\widetilde{\boldsymbol{x}}_{it}^{(\mathbf{0})})^\top$;

for *each episode* $m = 2, 3, \ldots$ **do**
- Compute $\tau_m \leftarrow 2^{m-\mathbf{1}}$, $\mathcal{T}_m \leftarrow \{\tau_{m-\mathbf{1}}+1, \ldots, \tau_m\}$;
- $\widehat{\boldsymbol{\nu}}_m \leftarrow$ `AggregateThenDebias`$(\mathcal{L}_t^{(h)}, \mathcal{L}_t^{(\mathbf{0})}, \lambda_m)$
- Set $W_{m-\mathbf{1}} \leftarrow V_{\tau_{m-\mathbf{1}}}^{(\mathbf{0})} + \sum_{h=\mathbf{1}}^{H} \omega_h V_{\tau_{m-\mathbf{1}}}^{(h)}$;
- Reset $V_{\tau_{m-\mathbf{1}}}^{(h)} \leftarrow 0_{\mathbf{2}d\times\mathbf{2}d}$, $\forall h \in \{0\} \cup [H]$;
- **for** *each period* $t \in \mathcal{T}_m$ **do**
  - $(S_t^{(\mathbf{0})}, \boldsymbol{p}_t^{(\mathbf{0})}) \leftarrow$ `OfferAssortmentAndPrice`$(t, \tau_m, q_m, V_t^{(\mathbf{0})}, \widehat{\boldsymbol{\nu}}_m, \alpha_m, \beta_m, W_{m-\mathbf{1}})$
  - **for** $h \in \{0\} \cup [H]$ **do**
    - $V_{t+\mathbf{1}}^{(h)} \leftarrow V_t^{(h)} + \sum_{i\in S_t^{(h)}} q_{it}^{(h)}(\widehat{\boldsymbol{\nu}}_m)\, \widetilde{\boldsymbol{x}}_{it}^{(h)} (\widetilde{\boldsymbol{x}}_{it}^{(h)})^\top -$
      $\sum_{i\in S_t^{(h)}} \sum_{j\in S_t^{(h)}} q_{it}^{(h)}(\widehat{\boldsymbol{\nu}}_m) q_{jt}^{(h)}(\widehat{\boldsymbol{\nu}}_m)\, \widetilde{\boldsymbol{x}}_{it}^{(h)} (\widetilde{\boldsymbol{x}}_{jt}^{(h)})^\top$;

Regret Upper Bound

THM. Regret Upper Bound (Informal)

Running Algorithm 3, the cumulative regret up to horizon T satisfies

$$\mathbb{E}\big[\,\mathrm{Regret}(T;\pi)\big] \;\le\; C_0\,\frac{\sqrt{KT}\,\log K}{L_0}\left(\sqrt{\frac{d\,\log T}{H+1}} + s_0\,\sqrt{\log(dT)}\right) \tag{6}$$

where $C_0, C_1 > 0$ are absolute constants depending only on $C_{\min}, C_{\max}$.

(i) Variance gain. With homogeneous covariates, self-normalized widths scale as $1/\sqrt{1+H}$ and yield the factor $\sqrt{1/(1+H)}$.
(i) Transfer bias. The second term comes from sparse target-only shifts; its radius scales with s_0. If $s_0{=}0$, the bias vanishes and we keep the full $1/\sqrt{1+H}$ speedup.
(iii) Baselines. Setting $H{=}0$ recovers the contextual rate. Compared to CAP, we match $\sqrt{KT}$ (up to logs) but improve dimension from d to $\sqrt{d}$ via sharper concentration.

Regret Lower Bound

THM. Regret Lower Bound (Informal)

For any $d \geq 1$, $K \in [d]$, and $s_0 \in \{0, 1, \ldots, \min\{K, d\}\}$, there exists a constant $c_2 = c_2(L_0, \overline{P}, C_{\min}, C_{\max}) > 0$ such that for all horizons T,

$$\inf_{\pi} \sup \mathbb{E}\big[\mathrm{Regret}(T; \pi)\big] \;\geq\; c_2 \left(\sqrt{\frac{K\,(d - s_0)\,T}{1 + H}} \;+\; s_0 \sqrt{K\,T} \right). \qquad (7)$$

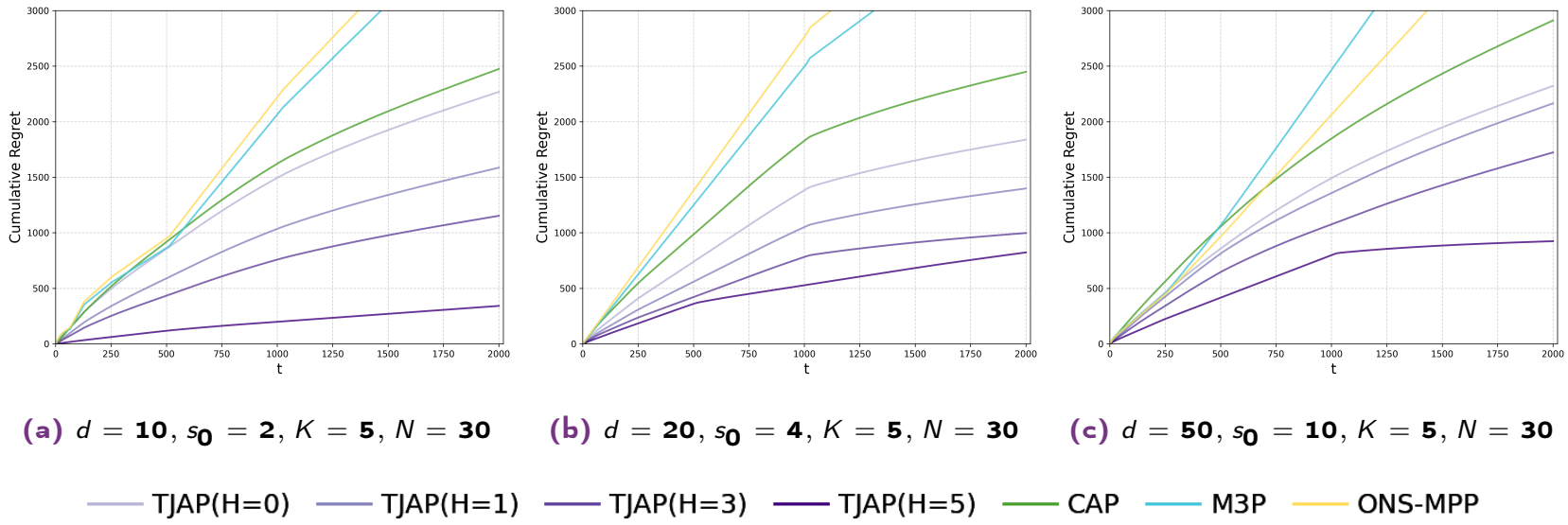
(i) Shared coordinates. The $\sqrt{K(d-s_0)T/(1+H)}$ term is the variance floor on the coordinates where sources and target agree; homogeneous covariates give the $1/\sqrt{1+H}$ gain.
(ii) Target-only coordinates. The $s_0\sqrt{KT}$ term reflects adaptation to coordinates unobserved in sources; no $1/\sqrt{1+H}$ improvement is possible there.
(iii) Tightness. Up to polylog factors and constants, the upper and lower bounds match in K, T, d, H, s_0.

Experimental Setup

Synthetic Data: auxiliary markets $H \in \{0, 1, 3, 5\}$, feature dimension $d \in \{10, 20, 50\}$, sparsity level $s_0 \in \{0.2d, 0.3d\}$, catalog and assortment capacity $(K, N) = (5, 30), (5, 100)$, time horizon $T = 2000$. A total of $3 \times 2 \times 2 = 12$ configurations.

Baselines and Protocol: *CAP* (Erginbas et al. 2025) is a contextual joint assortment–pricing algorithm, a direct comparison in the *no-transfer* specialization. *M3P* (Javanmard et al. 2020) and *ONS–MPP* (Perivier & Goyal 2022) are pricing-only methods; to place them in the JAP setting, at each period we rank items by the current utility estimate, select the top K as the assortment, and apply the method's posted prices to this subset.

Results & Findings



(a) $d = 10, s_0 = 2, K = 5, N = 30$ **(b)** $d = 20, s_0 = 4, K = 5, N = 30$ **(c)** $d = 50, s_0 = 10, K = 5, N = 30$

(i) Transfer helps. Within TJAP, regret decreases monotonically with the number of sources H.

(ii) CAP vs. TJAP. With transfer enabled, TJAP consistently outperforms CAP. Even in the *no-transfer* case ($H = 0$), TJAP remains stronger than CAP.

(iii) JAP vs. pricing-only. Even without transfer ($H = 0$), TJAP outperforms the pricing-only baselines (M3P and ONS–MPP).

Conclusion

Algorithmic: We propose a transfer learning framework for *joint* assortment–pricing that leverages source markets to accelerate learning, featuring *aggregate–then–debias* pipeline and *two-radius UCB*.

Theoretical: We establish *matching upper and lower regret bounds*, showing that transfer substantially accelerates learning when markets share exploitable structure.

Empirical: Simulations confirm *consistent regret reductions* compared with target-only baselines, underscoring the value of cross-market information for dynamic assortment–pricing.

Future work: Adaptive selection of informative sources and extending beyond ℓ_0-sparsity to richer relatedness notions, further broadening the reach of transfer learning in operational decision-making.

References I

Abbasi-Yadkori, Y., Pál, D. & Szepesvári, C. (2011), 'Improved algorithms for linear stochastic bandits', *Advances in neural information processing systems* **24**.

Agrawal, S., Avadhanula, V., Goyal, V. & Zeevi, A. (2017), Thompson sampling for the mnl-bandit, *in* 'Conference on learning theory', PMLR, pp. 76–78.

Agrawal, S., Avadhanula, V., Goyal, V. & Zeevi, A. (2018), 'Mnl-bandit: A dynamic learning approach to assortment selection'.
URL: *https://arxiv.org/abs/1706.03880*

Bastani, H. (2021), 'Predicting with proxies: Transfer learning in high dimension', *Management Science* **67**(5), 2964–2984.

Bastani, H., Simchi-Levi, D. & Zhu, R. (2022), 'Meta dynamic pricing: Transfer learning across experiments', *Management Science* **68**(3), 1865–1881.

Cai, Y., Cai, T. T. & Li, H. (2024), 'Transfer learning for contextual multi-armed bandits', *arXiv preprint* . v2, 2024.

Chen, W., Wang, Y. & Yuan, Y. (2013), Combinatorial multi-armed bandit: General framework and applications, *in* 'ICML', pp. 151–159.

Chen, X., Wang, Y. & Zhou, Y. (2020), 'Dynamic assortment optimization with changing contextual information', *Journal of machine learning research* **21**(216), 1–44.

Cheung, W. C. & Simchi-Levi, D. (2017), 'Thompson sampling for online personalized assortment optimization problems with multinomial logit choice models', *Available at SSRN 3075658* .

Dani, V., Hayes, T. P. & Kakade, S. M. (2008), Stochastic linear optimization under bandit feedback, *in* '21st Annual Conference on Learning Theory', number 101, pp. 355–366.

Erginbas, Y. E., Courtade, T. A. & Ramchandran, K. (2025), 'Online assortment and price optimization under contextual choice models', *arXiv preprint arXiv:2503.11819* .

Filippi, S., Cappe, O., Garivier, A. & Szepesvári, C. (2010), 'Parametric bandits: The generalized linear case', *Advances in neural information processing systems* **23**.

Javanmard, A., Nazerzadeh, H. & Shao, S. (2020), Multi-product dynamic pricing in high-dimensions with heterogeneous price sensitivity, *in* '2020 IEEE International Symposium on Information Theory (ISIT)', IEEE, pp. 2652–2657.

Kveton, B., Szepesvári, C., Wen, Z. & Ashkan, A. (2015), Cascading bandits: Learning to rank in the cascade model, *in* 'ICML', pp. 767–776.

References II

Li, L., Lu, Y. & Zhou, D. (2017), Provably optimal algorithms for generalized linear contextual bandits, *in* 'Proceedings of the 34th International Conference on Machine Learning - Volume 70', ICML'17, JMLR.org, p. 2071–2080.

Li, S., Cai, T. T. & Li, H. (2022), 'Transfer learning for high-dimensional linear regression: Prediction, estimation and minimax optimality', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **84**(1), 149–173.

Liu, S., Xu, A., Li, Z., Wu, J. & Fan, J. (2023), 'Unified transfer learning models for high-dimensional data', *arXiv preprint* .

Miao, S. & Chao, X. (2021), 'Dynamic joint assortment and pricing optimization with demand learning', *Manufacturing & Service Operations Management* **23**(2), 525–545.

Oh, M.-h. & Iyengar, G. (2019), *Thompson sampling for multinomial logit contextual bandits*, Curran Associates Inc., Red Hook, NY, USA.

Oh, M.-h. & Iyengar, G. (2021), Multinomial logit contextual bandits: Provable optimality and practicality, *in* 'Proceedings of the AAAI conference on artificial intelligence', Vol. 35, pp. 9205–9213.

Perivier, N. & Goyal, V. (2022), 'Dynamic pricing and assortment under a contextual mnl demand', *Advances in Neural Information Processing Systems* **35**, 3461–3474.

Rusmevichientong, P. & Tsitsiklis, J. N. (2010), 'Linearly parameterized bandits', *Mathematics of Operations Research* **35**(2), 395–411.

Tian, Y. & Feng, Y. (2022), 'Transfer learning under high-dimensional generalized linear models', *Journal of the American Statistical Association* **118**(544), 2684–2697.

Wang, R. (2012), 'Capacitated assortment and price optimization under the multinomial logit model', *Operations Research Letters* .

Xu, K. & Bastani, H. (2024), 'Multitask learning and bandits via robust statistics', *arXiv preprint* . Latest version accessed 2025.