Counterfactual Inference for Consumer Choice With Many Products

Susan Athey, Stanford University

Counterfactual Inference for Consumer Choice Across Many Product Categories (Susan Athey, David Blei, Rob Donnelly, Francisco Ruiz, in progress)
Consumer Choice
SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements (Francisco Ruiz, Susan Athey, David Blei, 2017)
Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data (Susan Athey, David Blei, Rob Donnelly, Francisco Ruiz, Tobias Schmidt, AEA Papers and Proceedings, 2018)

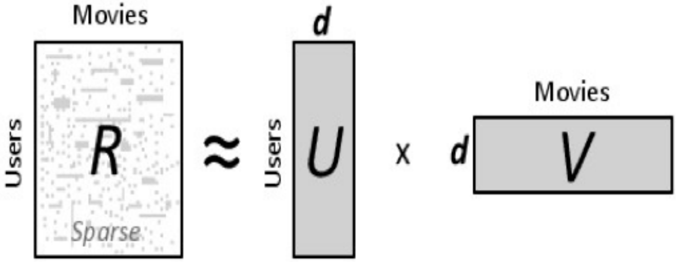
Asking and Answering Questions Using Panel Data with Consumer Choices Over Many Products

- Example Data Sets
  - List of Websites/Apps/News Articles Viewed
  - Device Id and Lat/Long of Locations Visited
  - Consumer Credit Card, Bank Transactions w/ Description
- Example Questions
  - How do users change their consumption of news when Google News shuts down, during election, when reading from Facebook, etc.?
  - How do physical movements of consumers change when they lose a job, when a store opens or closes?
  - How do consumers or suppliers in “gig economy” change spending patterns, travel, etc. as a result of the entry/increase in supply/changes in wages of Uber, Rover, etc.?
  - What products make good “loss leaders”? Interact w/ other products?
- Common features
  - Limited “structured” data about objects consumed
  - Consumers consumer wide variety of products, many rarely

Existing Approaches: ML

Typical approach from the machine learning literature:

- Canonical example: Netflix movie recommendations
- Estimating correlations in preferences between customers, ignoring substitutes/complements



"What types of things do customers like?"

Towards A Large Scale Model of Consumer Choice

One product at the time misses many aspects of consumer choice. E.g. for supermarkets:

- Store v. store competition happens at the level of a shopping trip, not an item
- Stores desire to understand profile of most valuable consumers, and attract them to the store
- Stores make decisions about products to stock and promote and how to price in order to attract different types of consumers
- Store organization can be made more or less convenient for different collections of products
- Bundling, loss leader strategies

Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data

Susan Athey, David Blei, Robert Donnelly, Francisco Ruiz and Tobias Schmidt

January 2018

Restaurant Choice

- Where should a restaurant be located?
- What is the best type of restaurant for a location?
- Who are a restaurant's competitors?
- How far will consumers travel to a restaurant they like?

These are examples of product design, location, and quality questions.

Travel Time Factorization Model (TTFM) of User Choice

$$U_{uit} = \underbrace{\lambda_i}_{\text{popularity}} + \underbrace{\theta_u^\top \alpha_i}_{\text{customer preferences}} - \underbrace{\gamma_u^\top \beta_i \cdot \log(d_{uit})}_{\text{distance effect}}$$
$$+ \underbrace{\mu_i^\top \delta_{w_{ut}}}_{\text{time-varying effect}} + \underbrace{\epsilon_{uit}}_{\text{noise}},$$

Covariates x_i affect mean of prior of α_i and β_i.

Logit error implies that choice probability conditional on going to a restaurant are:

$$Pr(Y_{ut} = i) = \frac{\exp \bar{U}_{uit}}{\sum_j \exp \bar{U}_{ujt}}$$

Travel Time Factorization Model (TTFM) of User Choice

$$U_{uit} = \underbrace{\lambda_i}_{\text{popularity}} + \underbrace{\theta_u^\top \alpha_i}_{\text{customer preferences}} - \underbrace{\gamma_u^\top \beta_i \cdot \log(d_{uit})}_{\text{distance effect}}$$
$$+ \underbrace{\mu_i^\top \delta_{w_{ut}}}_{\text{time-varying effect}} + \underbrace{\epsilon_{uit}}_{\text{noise}},$$

Covariates x_i affect mean of prior of α_i and β_i.

MNL Comparison: λ_i is constant across restaurants, α_i is observable characteristics of items, θ_u is constant across users, δ_w is omitted, and $\gamma_u \cdot \beta_i$ is constant across users and restaurants.

Dataset

Base Data

- SafeGraph, which aggregates locational information from consumers who have opted into sharing their location through mobile applications.
- "pings" from consumer phones: device id; timestamp; latitude, longitude
- January through October 2017, San Francisco Bay Area

Constructed Data

- "Typical" morning location of the consumer, defined as the most common place the consumer is found from 9:00 to 11:15 a.m. on weekdays.
- Most morning pings in morning location
- South San Francisco to San Jose, excl. mountains/coast
- Lunch restaurant visit: observed at least two pings more than 3 minutes apart during the hours of 11:30 a.m. to 1:30 p.m. in a location that we identify as a restaurant.
- Restaurants are identified using data from Yelp that includes geo-coordinates, star ratings, price range, restaurant

Summary Statistics

Table: Summary Statistics.

User-Level Statistics					
Variable (Per User)	Mean	25%	50%	75%	% Missing
Total Visits	11.63	4.00	7.00	13.00	—
Distinct Visited Rest.	7.25	3.00	5.00	9.00	—
Distinct Visited Categories	11.60	6.00	10.00	15.00	—
Median Distance (mi.)	3.06	0.89	1.86	3.79	—
Weekly Visits	0.39	0.15	0.25	0.47	—
Weeks Active	31.14	22.00	33.00	41.00	—
Mean Rating of Visited Rest.	3.29	3.00	3.33	3.61	1
Mean Price Range of Visited Rest.	1.55	1.33	1.53	1.75	0.6
Restaurant-Level Statistics					
Variable (Per Restaurant)	Mean	25%	50%	75%	% Missing
Distinct Visitors	13.53	5.00	10.00	19.00	—
Median Distance (mi.)	2.39	0.93	1.72	2.94	—
Weeks Open	42.17	44.00	44.00	44.00	—
Weekly Visits (Opens)	0.54	0.17	0.37	0.72	—
Weekly Visits (Always Open)	0.52	0.16	0.34	0.68	—
Weekly Visits (Closes)	0.53	0.15	0.34	0.67	—
Price Range	1.56	1.00	2.00	2.00	10.66
Rating	3.38	2.89	3.53	4.00	14.52

Estimation Details

- Bayesian Estimation with Hierarchical Prior
- Gaussian prior over latent char's, shifted by x_i:

$$p(\alpha_i \mid H_\alpha, x_i) = \frac{1}{(2\pi\sigma_\alpha^2)^{k_1/2}} \exp\left\{-\frac{1}{2\sigma_\alpha^2}||\alpha_i - H_\alpha x_i||_2^2\right\},$$
$$p(\beta_i \mid H_\beta, x_i) = \frac{1}{(2\pi\sigma_\beta^2)^{k_2/2}} \exp\left\{-\frac{1}{2\sigma_\beta^2}||\beta_i - H_\beta x_i||_2^2\right\}.$$

- Latent matrices H_α and H_β, of sizes $k_1 \times k_{\mathrm{obs}}$ and $k_2 \times k_{\mathrm{obs}}$ respectively, which weigh the contribution of each observed attribute on the latent attributes.
- Mean-field variational inference-approximate posterior with independent Gaussians and find parameters that minimize distance
- Stochastic gradient descent

Model Fit

Model	MSE	Log Likelihood	Precision@1	Precision@5	Precision@10
Training Sample					
TTFM	0.00025	-3.59	31.8%	59.4%	70.3%
MNL	0.00031	-6.58	2.8%	10.7%	16.7%
Held-out Test Sample					
TTFM	0.00028	-5.19	20.5%	35.5%	42.2%
MNL	0.00031	-6.55	3.1%	11.4%	17.5%

Figure: Predicted Versus Actual Shares By Distance

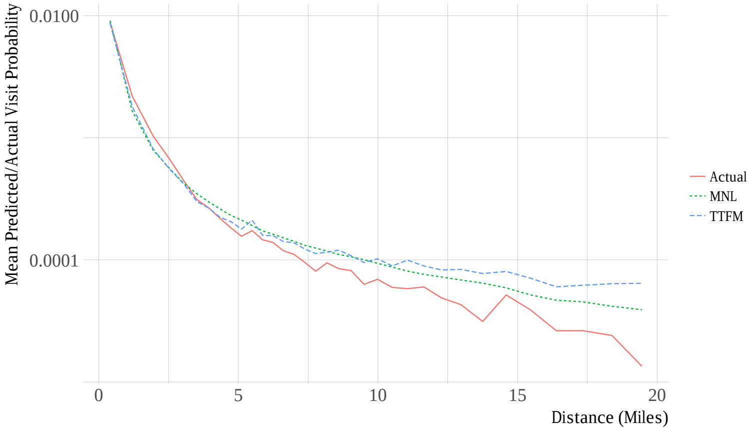


Figure: Actual v. Predicted Visits by Restaurant Visit Decile

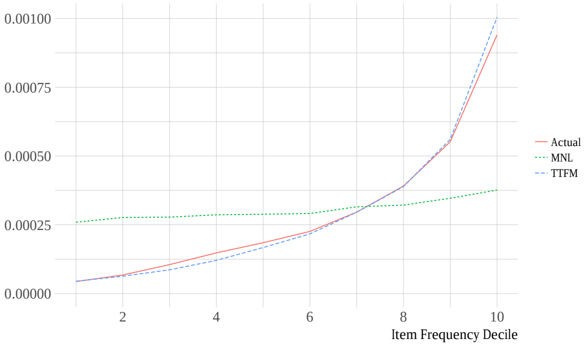


Table: Average Elasticities by Restaurant Characteristics, TTFM model.

Characteristic	Mean	se	25 %	50 %	75 %	N
All restaurants	-1.411	0.0001	-1.585	-1.408	-1.203	4924
Most popular category: Mexican	-1.499	0.0004	-1.664	-1.491	-1.285	694
Most popular category: Sandwiches	-1.435	0.0006	-1.602	-1.441	-1.235	522
Most popular category: Hotdog	-1.403	0.0007	-1.570	-1.390	-1.216	377
Most popular category: Coffee	-1.390	0.0008	-1.563	-1.404	-1.178	365
Most popular category: Bars	-1.370	0.0009	-1.546	-1.362	-1.161	352
Most popular category: Chinese	-1.353	0.0009	-1.517	-1.378	-1.176	350
Most popular category: Japanese	-1.320	0.0011	-1.472	-1.336	-1.140	276
Most popular category: Pizza	-1.497	0.0010	-1.649	-1.481	-1.307	260
Most popular category: Newamerican	-1.323	0.0019	-1.540	-1.351	-1.117	181
Most popular category: Vietnamese	-1.328	0.0020	-1.541	-1.327	-1.155	156
Most popular category: Other	-1.411	0.0002	-1.582	-1.406	-1.189	1391
Price range: 1	-1.446	0.0001	-1.607	-1.435	-1.245	2091
Price range: 2	-1.368	0.0001	-1.542	-1.371	-1.162	2165
Price range: 3	-1.320	0.0026	-1.506	-1.353	-1.108	122
Price range: 4	-1.449	0.0178	-1.664	-1.496	-1.289	21
Price range: missing	-1.474	0.0006	-1.648	-1.455	-1.225	525
Rating, quintile: 1	-1.427	0.0003	-1.605	-1.414	-1.209	842
Rating, quintile: 2	-1.392	0.0003	-1.557	-1.397	-1.187	842
Rating, quintile: 3	-1.364	0.0003	-1.532	-1.366	-1.169	842
Rating, quintile: 4	-1.385	0.0004	-1.571	-1.370	-1.180	842
Rating, quintile: 5	-1.438	0.0003	-1.603	-1.438	-1.250	841
Rating, quintile: missing	-1.475	0.0004	-1.653	-1.464	-1.232	715

Table: Average Elasticities by City, TTFM model.

Characteristic	Mean	se	25 %	50 %	75 %	N
All restaurants	-1.411	0.0001	-1.585	-1.408	-1.203	4924
City: Daly City	-1.105	0.0019	-1.331	-1.150	-0.959	165
City: Burlingame	-1.119	0.0030	-1.327	-1.194	-1.018	110
City: Millbrae	-1.130	0.0049	-1.418	-1.240	-0.954	80
City: San Bruno	-1.132	0.0035	-1.398	-1.216	-0.987	101
City: South San Francisco	-1.187	0.0021	-1.413	-1.232	-0.999	135
City: San Mateo	-1.243	0.0012	-1.454	-1.284	-1.101	268
City: Foster City	-1.318	0.0070	-1.506	-1.397	-1.163	44
City: San Carlos	-1.321	0.0026	-1.479	-1.350	-1.195	95
City: Palo Alto	-1.330	0.0013	-1.519	-1.342	-1.171	234
City: Brisbane	-1.332	0.0139	-1.455	-1.344	-1.181	15
City: Belmont	-1.334	0.0047	-1.500	-1.374	-1.212	58
City: Redwood City	-1.362	0.0012	-1.530	-1.389	-1.217	214
City: Cupertino	-1.365	0.0018	-1.532	-1.386	-1.174	169
City: East Palo Alto	-1.374	0.0142	-1.521	-1.393	-1.229	13
City: Los Gatos	-1.391	0.0026	-1.583	-1.437	-1.219	106
City: Los Altos	-1.406	0.0043	-1.564	-1.394	-1.236	60
City: Menlo Park	-1.407	0.0031	-1.570	-1.428	-1.287	87
City: Mountain View	-1.422	0.0013	-1.592	-1.429	-1.233	213
City: Santa Clara	-1.442	0.0009	-1.681	-1.456	-1.238	355
City: San Jose	-1.451	0.0002	-1.635	-1.464	-1.278	1858
City: Campbell	-1.482	0.0015	-1.640	-1.493	-1.317	144
City: Saratoga	-1.497	0.0059	-1.628	-1.481	-1.394	40

Figure: Model Predictions of the Effect of Restaurant Openings and Closings Controlling for Other Changes.

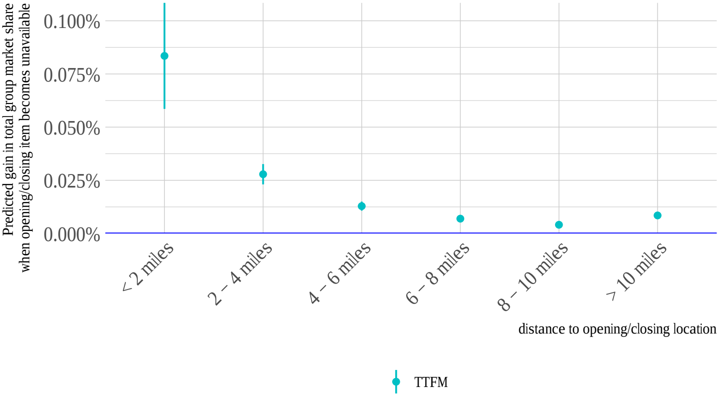


Table: Share of demand redistributed by distance, TTFM model

	Distance from opening/closing restaurant (mi.)					
	< 2	2 - 4	4 - 6	6 - 8	8 - 10	> 10
share	51 %	23 %	10 %	6 %	3 %	6 %
cum. share	51 %	74 %	84 %	90 %	94 %	100 %

Figure: Model Predictions Compared to Actual Outcomes for Restaurant Openings and Closings.

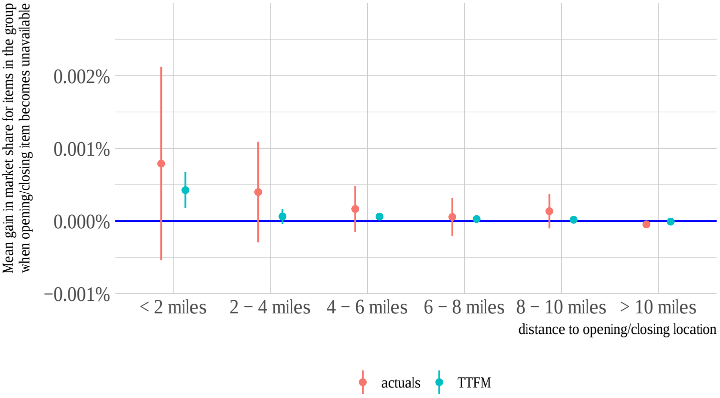
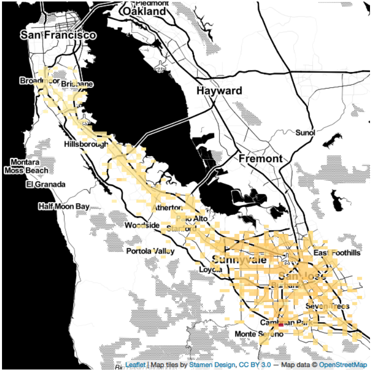


Figure: Best Locations for Restaurant Category

Cafes



Chicken Wings

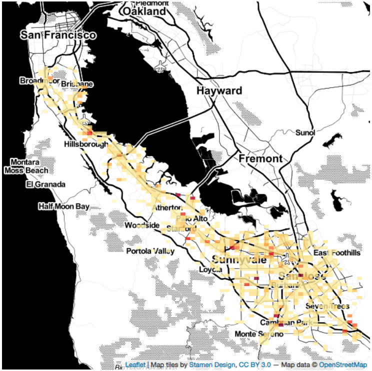
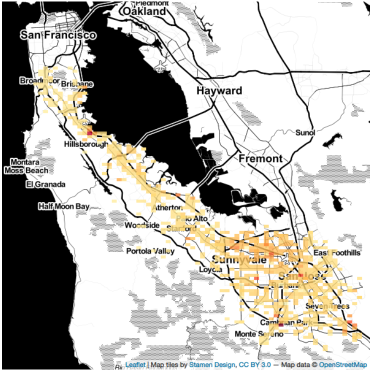


Figure: Best Locations for Restaurant Category

Filipino Restaurants



Sandwiches

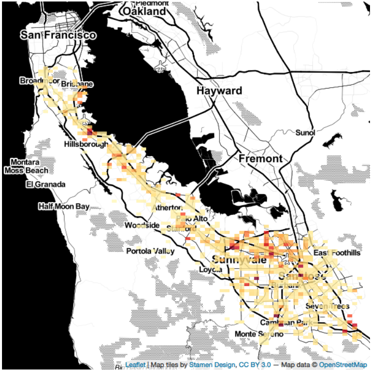
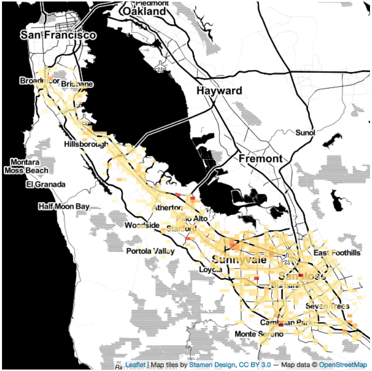


Figure: Best Locations for Restaurant Category

Vegetarian



Vietnamese Restaurants

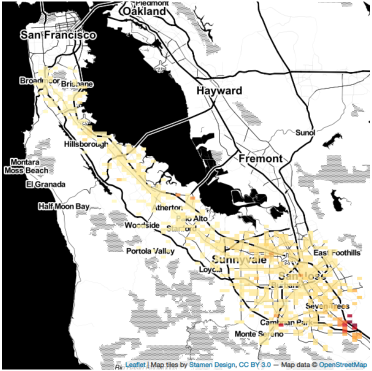
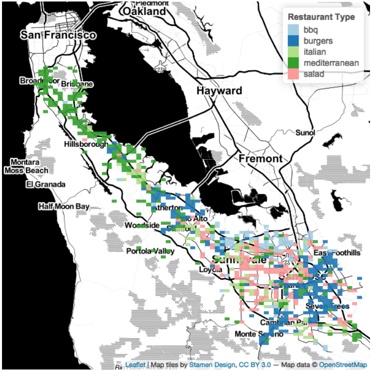


Figure: Best Restaurant Category for Locations

Mid-Priced ($$) Western Cuisine



Mid-Priced ($$) Asian Cuisine

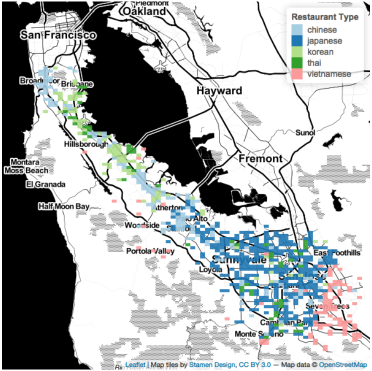
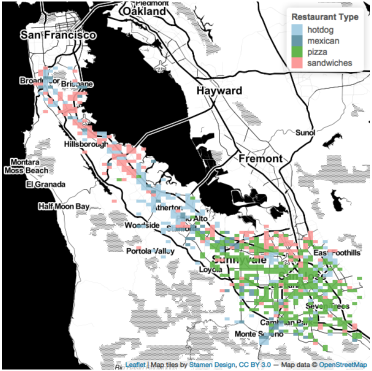
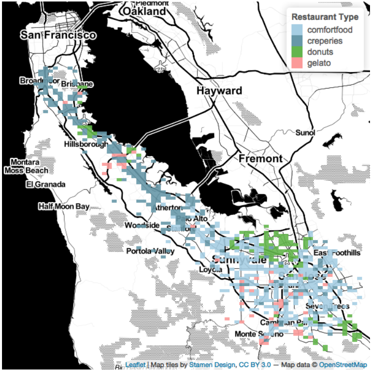


Figure: Best Restaurant Category for Locations

Cheap ($) Fast Food



Cheap ($) Treats



Economics/Marketing Literature Approaches to Demand

- Build a functional form model of user utility; estimate preferences based on user choice behavior
  - One product at the time (e.g. yoghurt)
  - Single product-specific latent variable (e.g. product quality) that may be correlated with price
  - Latent consumer preferences for observable product characteristics
  - Small number of papers: unobserved latent product characteristics (Goettler and Shachar, 2001; Athey-Imbens 2007; Nair, Misra, et al 2013)
- To identify effects of price
  - Instrumental variables approaches in cross-sectional data
  - Variety of approaches in panel data
- Other directions previously studied
  - Stockpiling, learning/experimentation, habit formation, effects of advertising/coupons/promotions

Asking and Answering Questions Using Panel Data with Consumer Choices Over Many Products

- Our Approach
  - Build on matrix factorization and embedding methods from CS
  - Use Bayesian approach for flexibility in incorporating structure
  - Estimate a structural model
  - Pay attention to identification and supplementary analyses in environment with many small experiments
- Two Ways to Use the Results
  - Traditional structural model: counterfactuals within the model
  - As a pre-processing step–to reduce the dimensionality and gain efficiency
    - Infer user preferences for rarely purchased products, rather than fixed effects
    - Motivate functional form assumptions–which products are potentially substitutes, complements, or independent
    - Reduce dimensionality of outcome space–e.g. Rover.com customers consumer more in latent categories related to travel

Towards A Large Scale Model of Consumer Choice

Goals of this agenda:

- Apply large-scale latent variable approaches with multiple unobserved product characteristics and latent user preferences over items (Poisson factorization: Gopalan, Hofman, Blei (2013); Gopalan, Charlan, Blei (2014))
- Consider many products at once
- Distinguish correlation from complementarity (price/availability change over time)
- Test assumptions and validate causal models
- Counterfactuals: consumer welfare for different pricing policies
- Evaluate the effects of policy changes, product introductions, or shocks to consumers

Steps

1. Validate price identification strategy systematically
2. Item by item choice: unified model v. separate by category
3. Complements

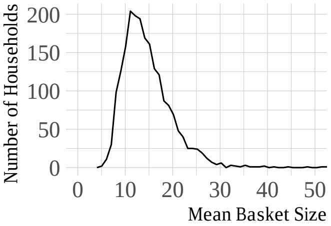
The Data

- Dataset constructed by Che, Chen and Chen (2012)
- All loyalty-card shoppers at a single, isolated store over 18 month period
- Out of stock data at approximately hourly level (daily is good enough)
- Almost all prices change on Tuesday night; focus on Tuesday and Wednesday data
- Product hierarchy (UPC, subclass, class, category, group, department section, department)
- User demographics: we include 28 variables derived from a variety of sources, may have measurement error

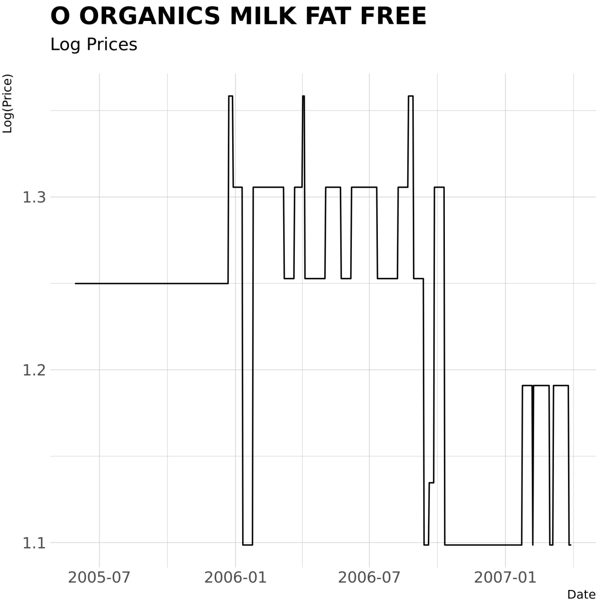
The Data: Sample Selection

- Users who had at least 20 shopping trips on Tuesday or Wednesday with more than 10 items per trip
- Top 235 categories
- Further restrictions
  - More than one UPC in category
  - More than one top ten product w/ price variation
  - Purchases not too concentrated
  - Prices not too highly correlated
  - Less than 10% buy multiple items per purchase in category
  - More than one top 10 UPC with price changes greater than $.10 in at least 10% of weeks
- Dataset: 2068 consumers, 123 categories, 1263 items, 333,585 trips
- Training/Validation/Test: 65%, 5%, 30%

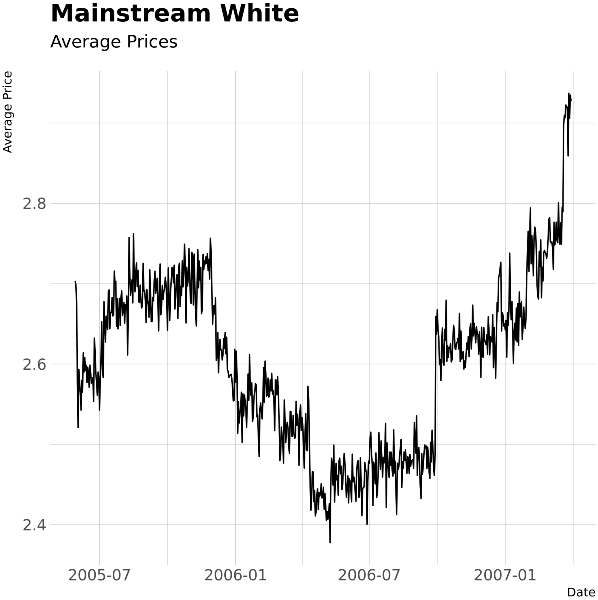
Shopping Size and Frequency



A Sample UPC-Level Price Series



A Sample Category-Level Price Series



Assumptions: Discussion

Our "identifying assumption" is:

- Counterfactual distributions of purchases between Tues & Wed in weeks with price changes can be constructed from behavior on weeks without price changes
- Weeks with price changes, and magnitudes of price changes, have the same day of week/time of day patterns as other weeks

This assumption would be violated if, for example:

- Store chooses to lower prices in a week where demand would have been growing through the week (e.g. due to minor holiday at end of week)
- Prices generally trending up or down in the sample, corresponding to decline or increase in a product's popularity, marketing, fruit coming into season, etc.

Testing Assumptions: Pseudo-Treatments

If prices and quantities both have systematic time trends, fixed effect models will not fully control for them

- Include global time trends as controls (no effect in most products)
- Placebo test
    - Shift price series forward (or backward) in time
    - Skip over weeks that also have a price change
    - Re-estimate model, calculate p-values

Supplementary Analysis

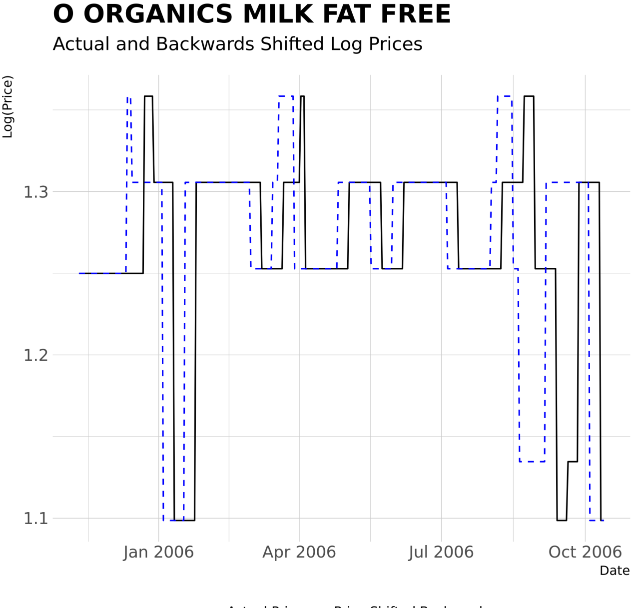
In simple model, do variants of "Placebo tests" to see if our identification strategy is valid.

- Use only shopping trips t on Tuesday and Wednesday
- Estimate multinomial logit model with

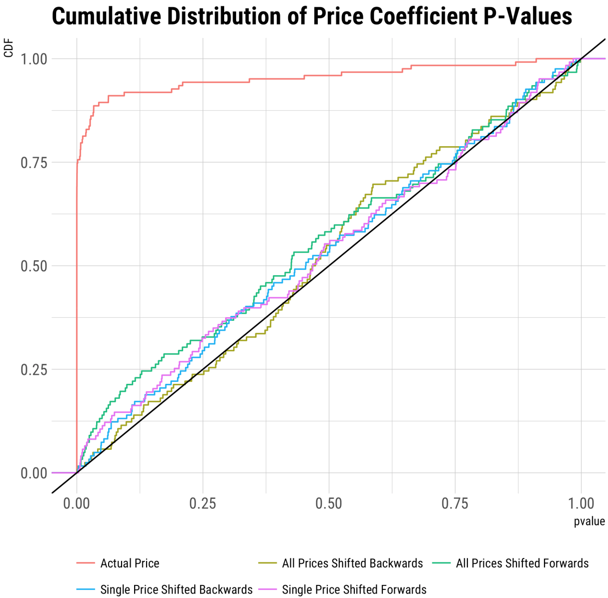
$$U_{uit} = \alpha \log(P_{it}) + X_{it}\beta' + \epsilon_{uit} \tag{1}$$

- $\log(P_{it})$ log price of item during trip
- X_{it} contains
  - Tuesday and Wednesday dummies
  - Week pseudo-fixed effects (week average category purchases)

A Sample UPC-Level Price Series



Placebo Test Results



The Baseline Poisson Factorization (HPF) Model

- User u has K-vector of non-negative preferences θ_u
- Item i has a K-vector of non-negative attributes β_i.
- Mean utility for item

$$\mu_{uit} = \log(\theta_u^\top \beta_i) \tag{2}$$

  where $U_{uit} = \mu_{uit} + \epsilon_{uit}$ and ϵ_{uit} is drawn from extreme value distribution.
- She chooses to buy if utility is positive.
- Choices about products are independent
- Parameters are all positive

The Nested Logit Factorization Model

- Choices in one category are independent of other categories
- User u has K-vector of preferences θ_u and a vector of price sensitivity parameters γ_u
- Item i has two K-vector of attributes α_i and β_i.
- Mean utility for item

$$\mu_{uit} = \theta_u^\top \beta_i - \gamma_u^\top \alpha_i p_{uit} \tag{3}$$

where $U_{uit} = \mu_{uit} + \epsilon_{uit}$ and ϵ_{uit} is drawn from extreme value distribution and are independent conditional on purchasing an item within a category, implying

$$Pr(Y_{ut} = i | \text{purchase in cat for } ut) = \frac{\exp \mu_{uit}}{\sum_{j>0} \exp \mu_{ujt}} \tag{4}$$

.

- She chooses the highest utility item in each category or the outside option; the outside option is in its own nest

The Nested Logit Factorization Model Cont'd

- Users u independently chooses whether or not to make a purchase from each product category c.
- Utility for not choosing category

$$U_{uc_0t} = \theta_{c,u}^{\top}\beta_{c_0} + \epsilon_{uc_0t} \tag{5}$$

- Utility for choosing category

$$U_{uc_1t} = \theta_{c,u}^{\top}\beta_{c_1} - \mu_u\delta_{c_1}IV_c + \epsilon_{uc_1t} \tag{6}$$

- Where IV_c is the inclusive value of the items in the category is given by $IV_c = \log\sum_{i\in J_c}\exp U_{uit}$, which is the expectation of the max of the U_{uit} prior to learning the ϵ_{uit} for each item.

Estimation of Model

- MCMC-based Bayesian methods: Common in marketing for estimating models with heterogeneity, but computationally infeasible as data size and number of parameters grows
- This choice of functional form allows for fast and efficient estimation using variational Bayesian inference
- Variational Bayes:
  - Choose parameterized family of distributions $q(\cdot|\eta)$ to approximate the posterior
  - Find η that minimizes KL-divergence to the true posterior
  - With appropriate choice of priors and q, this optimization can be done using simple coordinate ascent
  - Accuracy similar to MCMC, but 1000s of times more quickly
- Introducing price effects and time-varying price slows things down substantially (hours rather than minutes; but still feasible unlike MCMC)
- Introducing substitutability within categories requires additional computational tricks

Model Details

- Item characteristics: Department section indicator variables, Price
- Customer characteristics: Gender, Age Bracket, Marital Status, Children, Income Bracket
- Control for weekly average purchases at the category level
- Number of latent characteristics chosen through validation

Model Comparisons

- Main benefits of our model: personalization and efficiency from pooling categories
- Can compare to standard models (MNL, nested logit, mixed logit)
- Can also look at benefit from taking HPF estimates of personalized mean utility for each item, and including these in standard models
- Note bias-variance tradeoff issues:
  - If you evaluate at individual level, better to have model with high personalization but some bias; at aggregate level, need to eliminate bias with simpler model that gets average right

Model	Mean Log Likelihood		Mean Squared Error	
	Train	Test	Train	Test
Nested Factorization	-4.9096	-4.2271	0.9268	0.8981
Mixed Logit with Random Price and HPF Controls	-5.3125	-4.9233	0.9660	0.9473
Multinomial Logit with Item-Specific HPF	-5.4213	-5.2286	0.9648	0.9580
Nested Logit with HPF Controls	-5.4230	-5.2345	0.9650	0.9583
Multinomial Logit with HPF Controls	-5.4248	-5.2307	0.9651	0.9583
Hierarchical Poisson Factorization (HPF)	-5.4484	-5.2377	0.9604	0.9531
Mixed Logit with Random Price and Demographics	-5.5690	-5.2976	0.9898	0.9780
Mixed Logit with Random Price and Behavioral Controls	-5.5820	-5.3164	0.9905	0.9786
Mixed Logit with Random Price and Random Intercepts	-5.5827	-5.3956	0.9849	0.9785
Nested Logit with Demographic Controls	-5.6779	-5.6080	0.9801	0.9788
Multinomial Logit with Demographic Controls	-5.6791	-5.6142	0.9803	0.9791
Nested Logit with Behavioral Controls	-5.6933	-5.6303	0.9805	0.9794
Multinomial Logit with Behavioral Controls	-5.6939	-5.6349	0.9806	0.9796
Multinomial Logit	-5.7156	-5.6747	0.9810	0.9802
Nested Logit	-5.7180	-5.6722	0.9809	0.9800

Table: Comparison of Predictive Fit

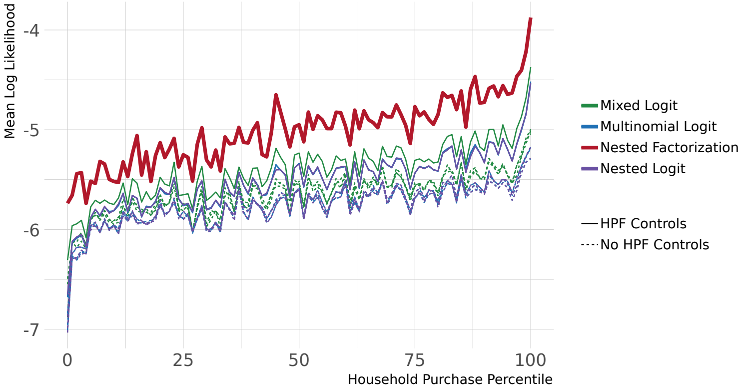


Figure: Comparison of Predictive Fit by Household Frequency

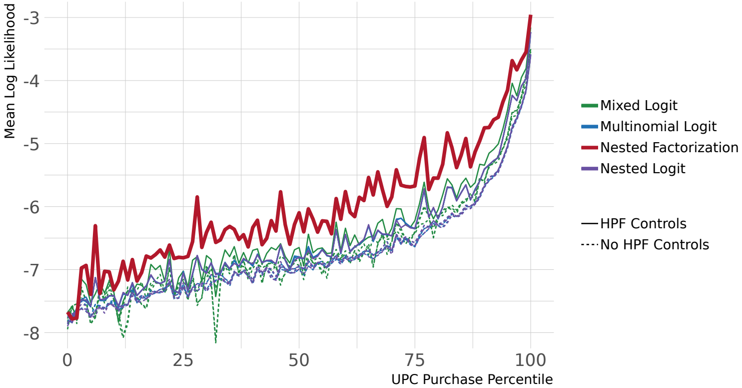


Figure: Comparison of Predictive Fit by UPC Frequency

Counterfactuals: Individual Level, Popular products

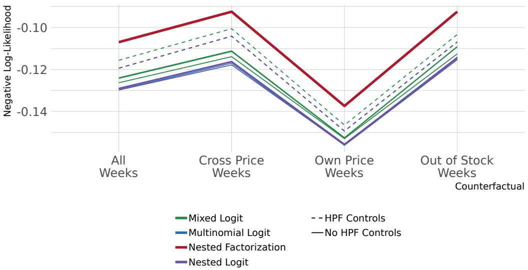


Figure: Mean log likelihood at the level of individual shopping trips is calculated as the average over all consumer shopping trips that occur during a week with the corresponding counterfactual event for the focal product. Popular products are defined as items that are purchased at least 2.5 times per day on average.

Counterfactuals: Individual Level, Unpopular Products

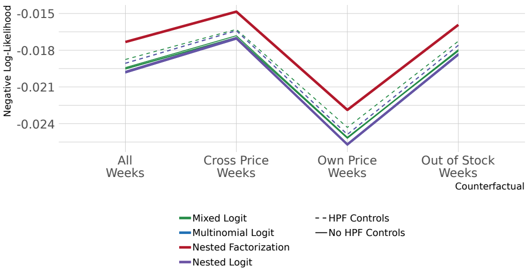


Figure: Mean log likelihood at the level of individual shopping trips is calculated as the average over all consumer shopping trips that occur during a week with the corresponding counterfactual event for the focal product. Popular products are defined as items that are purchased at least 2.5 times per day on average.

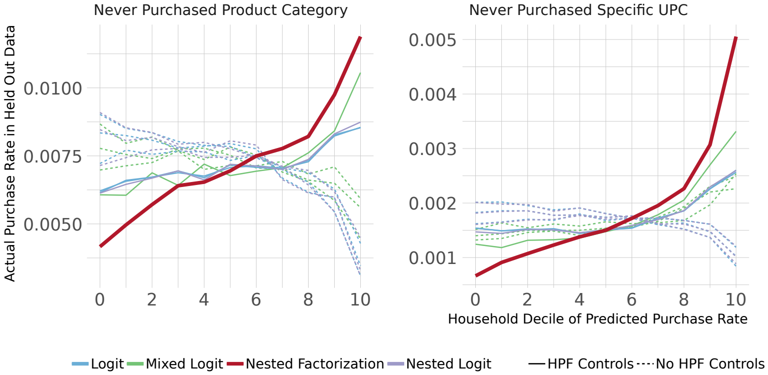


Figure: Test set purchase rate of households who made no purchases of the corresponding UPC/Category during the training data.

	Coef of Variation		Regression Coef	
Model	UPC	Category	UPC	Category
Nested Factorization	3.2546	1.7756	0.9955	1.0023
Mixed Logit with Random Price and HPF Controls	2.0747	1.6085	0.6861	0.7007
Mixed Logit with Random Price and Behavioral Controls	1.4049	1.6516	0.4575	0.5980
Mixed Logit with Random Price and Demographics	1.3869	1.5724	0.4718	0.5968
Multinomial Logit with HPF Controls	1.2590	0.7276	0.8402	0.8893
Nested Logit with HPF Controls	1.2368	0.7520	0.8417	0.8725
Multinomial Logit with Item-Specific HPF	1.2276	0.7318	0.8413	0.8881
Hierarchical Poisson Factorization (HPF)	1.1225	0.9345	0.9496	0.9256
Mixed Logit with Random Price and Random Intercepts	1.0834	1.0446	0.4666	0.6959
Nested Logit with Demographic Controls	0.4465	0.2967	0.8947	0.9314
Multinomial Logit with Demographic Controls	0.4337	0.2756	0.9077	0.9411
Nested Logit with Behavioral Controls	0.3539	0.2626	0.9524	0.9666
Multinomial Logit with Behavioral Controls	0.3507	0.2647	0.9527	0.9603
Nested Logit	0.1301	0.1172	1.0520	0.9213
Multinomial Logit	0.1236	0.1225	1.0655	0.8749

Table: Comparison of Degree of Personalization of Predictions across Models

The coefficient of variation is defined as $\frac{\text{sd(prediction)}}{\text{mean(prediction)}}$. Calculate this separately for each UPC / Category; report average.

The regression coefficient is a regression of the actual purchase rate on the predicted purchase rate with fixed effects at the UPC / Category

Model	Own Price			Class Cross Price			Subclass Cross Price		
	Median	SD(Mean)	Mean(SD)	Inside	Outside	%	Inside	Outside	%
Nested Factorization	-1.7121	1.2008	1.7774	0.0186	0.0080	132%	0.0196	0.0181	8.4%
Nested Logit with Demographic Controls	-1.2976	1.0377	0.0532	0.0119	0.0086	37%	0.0125	0.0115	8.5%
Nested Logit with Behavioral Controls	-1.3333	0.9791	0.0415	0.0129	0.0094	37%	0.0130	0.0129	1.0%
Nested Logit	-1.3298	0.9648	0.0105	0.0131	0.0096	37%	0.0131	0.0131	-0.2%
Multinomial Logit	-1.1864	0.8824	0.0003	0.0034	0.0029	17%	0.0034	0.0034	1.5%
Mixed Logit with Random Price and Random Intercepts	-2.2024	1.7822	0.8387	0.0062	0.0053	16%	0.0062	0.0062	-0.7%
Mixed Logit with Random Price and HPF Controls	-2.7077	1.9084	1.1294	0.0063	0.0054	16%	0.0063	0.0065	-3.9%
Multinomial Logit with HPF Controls	-1.1841	0.8904	0.0060	0.0035	0.0030	16%	0.0036	0.0035	0.8%
Multinomial Logit with Item-Specific HPF	-1.1850	0.8908	0.0060	0.0035	0.0030	16%	0.0036	0.0035	0.8%
Multinomial Logit with Demographic Controls	-1.1813	0.8837	0.0017	0.0034	0.0029	16%	0.0034	0.0034	2.1%
Multinomial Logit with Behavioral Controls	-1.1871	0.8835	0.0014	0.0032	0.0028	14%	0.0032	0.0032	0.6%
Mixed Logit with Random Price and Behavioral Controls	-3.0241	2.4619	1.4699	0.0064	0.0056	14%	0.0063	0.0066	-4.4%
Mixed Logit with Random Price and Demographics	-3.0897	2.4546	1.4785	0.0067	0.0060	13%	0.0066	0.0069	-4.6%
Nested Logit with HPF Controls	-1.1017	0.8783	0.0182	0.0148	0.0158	-7%	0.0140	0.0157	-10.7%

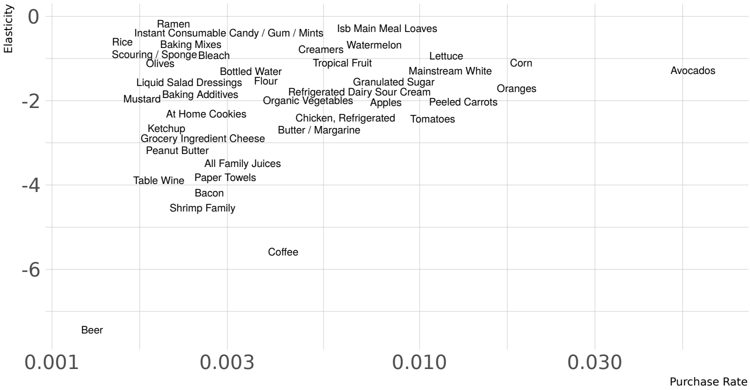


Figure: Category Level Elasticities and Predicted Purchase Probabilities:

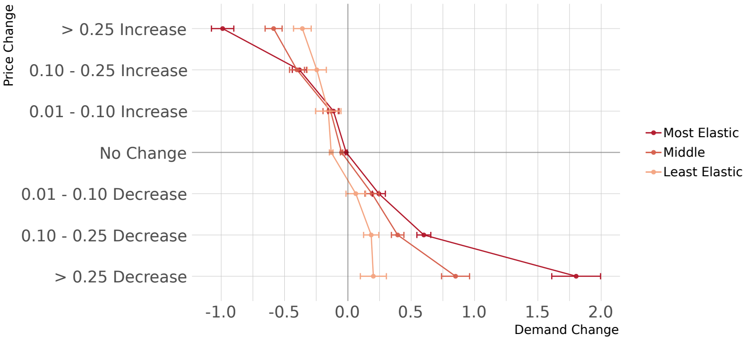


Figure: We calculate the mean change in aggregate demand between Tuesday and Wednesday as a function of the size of the price change. Consumers are split into 3 groups for each UPC based on their estimated elasticity.

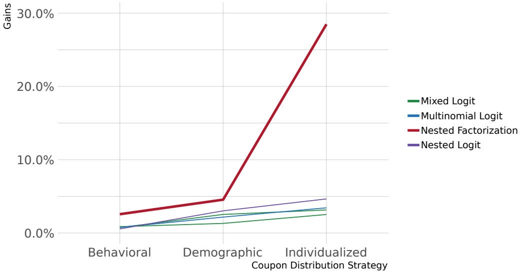
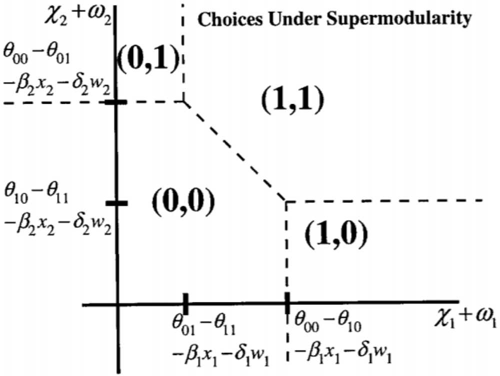


Figure: Gains from Targeted Discounts:
We evaluate the ability of the store to target a 30% coupon to 30% of it's customers based on each model's estimates, using the Nested Factorization model as ground truth. We compare the percentage gains in profits of different approaches to targeting these coupons relative to the profits of distributing the coupons uniformly at random.

Substitutes and Complements

FIGURE 1



Athey-Stern discuss identification through variation in X, W (like prices), and estimation through numerical integration of the unobservables

Substitutes and Complements

- Surprisingly little literature on empirical estimation since Athey and Stern (1998), and almost all have 2-3 choices
- Gentzkow (2007) implements similar approach
- Berry et al (2014), Chintagunta and Nair (2011) survey
- Train, McFadden and Ben-Akiva (1987) treat each bundle as a discrete alternative, but use nested logit to account for correlation among related bundles
- Song and Chintagunta (2007) build a utility-maximization framework where consumers select not just whether to purchase, but how much, and apply it to supermarket purchase data for two products, laundry detergent and fabric softener.
- See also Seiler, Thomassen, Smith and Schiraldi (2017); Smith, Rossi, and Allenby (2017); Wan, Wang, Goldman, Taddy, Rao (2017)

Consumer Shopping Heuristic

Behavioral assumption

- Assume myopic consumer goes into the store and considers sequentially what to buy
- Selects the item that maximizes utility myopically (for now: over whole store)–assuming will not buy anything else–but accounts for complementarity with what is already in basket
- "Look-ahead model": look ahead to the next item (for now: over whole store) when making choice, accounting for potential complementarity

Estimation: latent item orderings

- For each order, assume analyst observed that order
- Model implies a likelihood over each ordering (higher value items purchased earlier)

Simulation
Comparing Myopic to Look-Ahead Models

		stage 1: Diapers	stage 2: Hotdogs	stage 3: Buns	stage 4: *checkout*
non think-ahead	diapers	**0.39**	0.00	0.00	0.00
	coffee (↑)	0.01	0.02	0.03	0.12
	ramen	0.00	0.00	0.00	0.00
	candy	0.00	0.00	0.00	0.00
	hot dogs	0.18	**0.20**	0.00	0.00
	hot dog buns	0.18	0.24	**0.79**	0.00
	taco shells (↑)	0.03	0.07	0.00	0.00
	taco seasoning	0.21	0.41	0.00	0.00
	checkout	0.00	0.06	0.18	**0.88**
think-ahead	diapers	**0.38**	0.00	0.00	0.00
	coffee (↑)	0.02	0.01	0.06	0.05
	ramen	0.00	0.00	0.00	0.00
	candy	0.00	0.00	0.00	0.00
	hot dogs	0.23	**0.40**	0.00	0.00
	hot dog buns	0.32	0.53	**0.87**	0.00
	taco shells (↑)	0.02	0.02	0.00	0.00
	taco seasoning	0.02	0.03	0.00	0.00
	checkout	0.00	0.01	0.06	**0.95**

Goodness of Fit

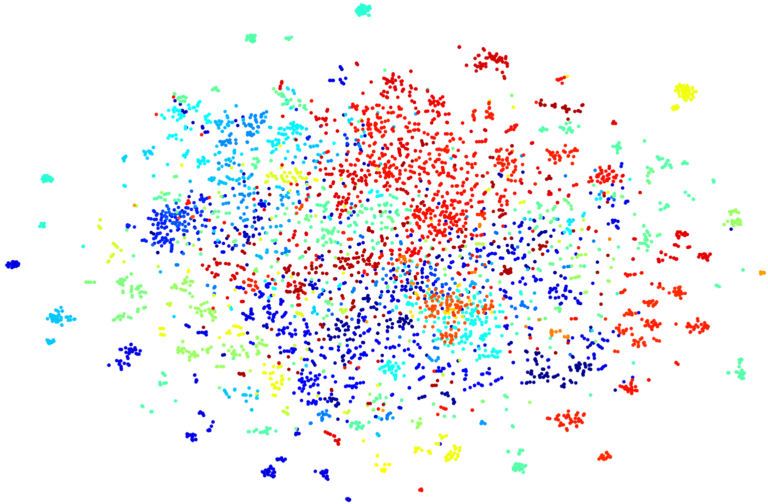
Model	Log-likelihood All (320K)	Price±10% (66K)	Price±20% (20K)	Price±30% (1.5K)
B-Emb	−5.13	−5.30	−5.33	−5.35
P-Emb	−5.13	−5.34	−5.42	−5.48
HPF	−4.97	−5.24	−5.35	−5.45
This paper (I+U)	−4.94	−5.21	−5.27	−5.33
This paper (I+U+P)	−4.93	−5.13	−5.09	−5.01
This paper (I+U+P+S)	**−4.92**	**−5.12**	**−5.08**	**−5.00**

Table: Average predictive log-likelihood on the test set, conditioning on the remaining items of each basket. SHOPPER with user preferences improves over the existing models. The improvement grows when adjusting for price and seasonal effects, and especially so when using skewed test sets that emulate price intervention.

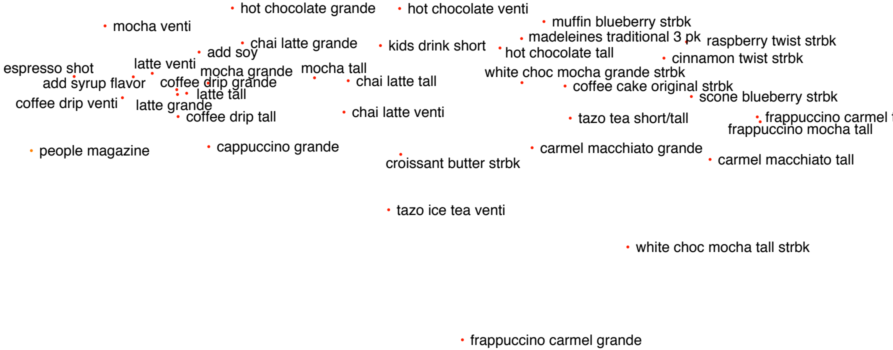
Role of Lookahead Model

	Three items	Entire baskets
Non think-ahead	-4.93	-5.14
Think-ahead	$-\mathbf{4.82}$	$-\mathbf{5.05}$

Representation of Latent Item Characteristics

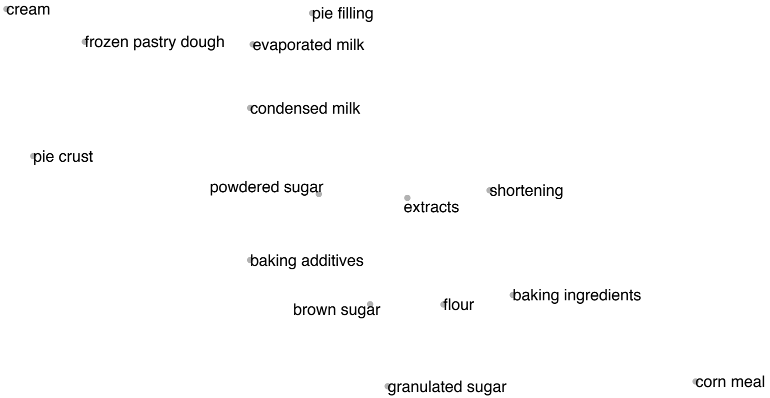


UPCs that are Close in α space

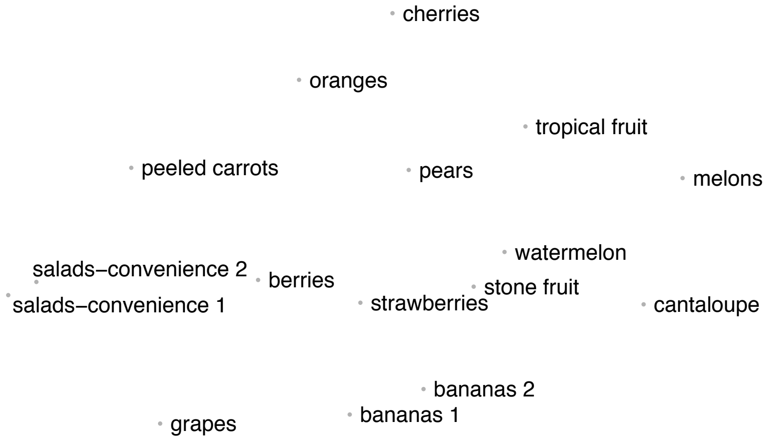


UPCs that are Close in α space

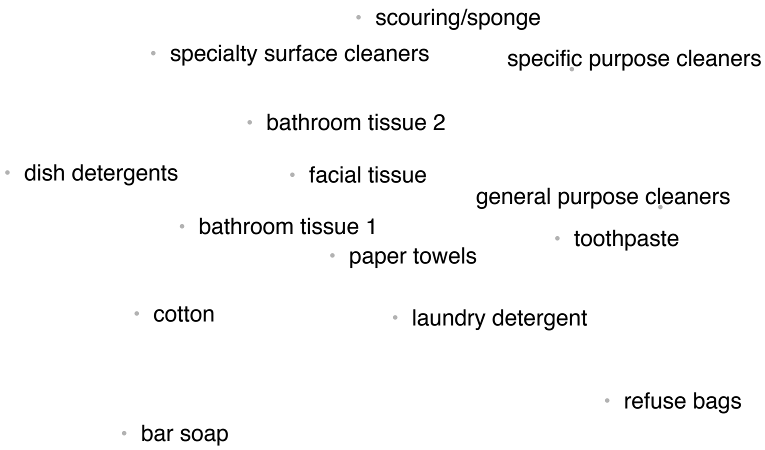
Categories that are Close in α space



Categories that are Close in α space



Categories that are Close in α space



Similarity/Exchangeability v. Complementarity

query items	complementarity score		exchangeability score	
mission tortilla soft taco	2.51	ortega taco shells white corn	0.05	mission fajita size
	2.40	mcrmck seasoning mix taco	0.10	mission tortilla fluffy gordita
	2.26	lawrys taco seasoning mix	0.11	mission tortilla soft taco
private brand hot dog buns	3.02	bp franks bun size	0.10	private brand hamburger buns
	2.94	bp franks beef bun length	0.12	ball park buns hot dog
	2.86	private brand hamburger buns	0.14	private brand hot dog buns ssme 8ct
private brand mustard squeeze bottle	0.53	private brand hamburger buns	0.14	frenchs mustard classic yellow squeeze
	0.44	private brand cutlery full size asst	0.16	frenchs mustard classic yellow squeezed
	0.29	private brand hot dog buns	0.17	heinz ketchup squeeze bottle
private brand napkins all occasion	1.01	private brand cutlery full size forks	0.08	vnty fair napkins all occasion
	0.62	dixie heavy duty plates dspbl 10 1/4 in	0.10	vnty fair napkins all occasion
	0.39	private brand plate dsgnr 6 7/8 in	0.13	glad cling wrap plastic wrap

Table: Items with the highest complementarity and lowest exchangeability metrics for some query items.

Conclusions

- Factorization is effective technique that enables personalization with dimension reduction
- Bayesian/structural models enable incorporation of functional forms motivated by theory and practice in economics
- Important to tune models for desired goals (counterfactuals)
- Models can discover product interaction and price sensitivities with minimal external information about products
- These models are a baseline against which we can proceed to estimate impact of other types of interventions