What can we learn from second choice data alone?

Chris Conlon (NYU Stern) and Julie Holland Mortimer (BC) and Paul Sarkis (BC)

May 1, 2021

IIOC 2021

Motivation

Diversion Ratios

The diversion ratio is one of the best ways we have measure competition between sellers.

- Raise the price of j and count the number of consumers who leave
- Diversion ratio is the fraction of leavers who switch to the substitute k.
- Useful because it enters the multi-product Bertrand FOC:

$$p_j(p_{-j}) = \underbrace{\frac{1}{1 + 1/\epsilon_{jj}(\mathbf{p})}}_{\text{Markup}} \left[c_j + \sum_{k \in \mathcal{J}_f \setminus j} (p_k - c_k) \cdot D_{jk}(\mathbf{p}) \right].$$

- A higher value of D_{jk} indicates closer substitutes.
- Conlon and Mortimer (2021) show second-choices and small price changes measure different averages.

What is our paper about?

We consider a problem where we observe some aggregate shares $\mathbf{s} = [s_1, \ldots, s_J]$ and some elements $(j,k) \in$ OBS of $\mathcal{D}$ a matrix of (second-choice) diversion ratios.

$$\mathcal{D} = \begin{array}{cc} \begin{array}{ccccc} \text{VZ} & \text{ATT} & \text{TMo} & \text{S} & \text{Other} \end{array} & \\ \begin{pmatrix} 0 & ? & ? & ? & ? \\ ? & 0 & ? & ? & ? \\ 0.3 & 0.45 & 0 & 0.2 & 0.05 \\ 0.3 & 0.15 & 0.45 & 0 & 0.1 \\ `? & ? & ? & ? & 0 \end{pmatrix} & \begin{array}{c} \text{VZ} \\ \text{ATT} \\ \text{TMo} \\ \text{S} \\ \text{Other} \end{array} \end{array}, \quad \begin{bmatrix} 0.35 \\ 0.30 \\ 0.20 \\ 0.15 \\ 0.05 \end{bmatrix} = \mathbf{s}$$

Can we fill in the missing elements?

How do we fill in missing elements?

Typical Approach: estimate a parametric model.

- Multi-product demand with unrestricted matrices of $(J+1)^2$ cross-elasticities (such as AIDS) is often hopeless with large J. Unrestricted diversion likely equally hopeless.
- Plain logit places strong restrictions: $D_{jk} = \frac{s_k}{1-s_j}$.
- Nested logit $D_{jk} = \frac{s_{k|g}}{Z(\sigma, s_g) - s_{j|g}}$ (same nest) where σ is nesting parameter.
- Mixed Logit: Explain substitution patterns using observed characteristics
  - Typically assume independent normal RC
  - Two products with similar x_1 and high substitution $\rightarrow$ larger σ_1.
  - Two products with similar x_2 and low substitution $\rightarrow$ smaller σ_2.
- McFadden and Train (2000) show some mixed logit is fully flexible
  - And a sufficient set of characteristics x to explain $\mathcal{D}$
  - But this depends on $f(\beta_i)$ heterogeneity being nonparametric

How do we fill in missing elements?

Our paper: Consider a low-rank approximation to $\mathcal{D}$

- Limit the rank of $\mathcal{D}$ directly in product space without parametric restrictions.
- Instead of controlling complexity with product characteristics and parametric restrictions on random coefficients.
- Allow for sparsity in substitution patterns (could be limited consideration).
- Much better at generating extreme patterns for best substitutes.

Why? This works well in other domains (CS for image recovery/compression).
We show this actually has a sensible economic interpretation!

Why might we want to do this?

There may be lots of cases where:

- We have access to aggregate market shares and some (but not all) second choice data
- We are interested in estimating substitution patterns across all sets of products
  - e.g. If we data on shares of largest cellular phone providers, and number porting or switching data for merging parties only.
  - e.g. We have survey data on "If this Tesco were to close where would you shop" (as UK CMA likes to ask).
  - e.g. We have other win-loss data for merging parties only.
- In many cases we may not have sufficient variation in prices (or other covariates) to estimate a demand system.
- Product characteristics may not accurately capture substitution across products.

Setup and Model

Recap of Conlon Mortimer (RJE conditionally accepted)

Considers the Diversion Ratio as a Local Average Treament Effects (LATE) estimator:

$$\frac{q_k(z_j', x) - q_k(z_j, x)}{-\Big(q_j(z_j', x) - q_j(z_j, x)\Big)} = \mathbb{E}[D_{jk,i}(x) | d_{ij}(z_j, x) > d_{ij}(z_j', x)]$$

- For discrete choice models we can measure diversion using any intervention z_j (price, quality, other characteristics).
- Interventions measure the average diversion ratio among a set of compliers (people who leave as we change z_j).
- Second-choice data considers the case where everyone is treated (nobody buys j): $\mathbb{E}[D_{jk,i}(x) | d_{ij}(z_j', x) = 0]$.

Recap of Conlon Mortimer (RJE conditionally accepted)

We showed that any average diversion ratio can be decomposed into:

$$\int D_{jk,i}(x) \cdot w_i(z_j, z_j', x) \partial F_i$$

- Individual diversion ratios $D_{jk,i}(x)$ don't depend on the intervention.
- Weights $w_i(z_j, z_j', x)$ determine who leaves j but aren't related to the substitute k.
- For the mixed logit: $D_{jk,i}(x) = \frac{s_{ik}(x)}{1-s_{ij}(x)}$ (proportional substitution for each type i).
- For second-choice data $w_{ij}(x) = \frac{s_{ij}(x)}{s_j(x)}$ (fraction of type i among j buyers).

For any mixed logit the expression for second choice diversion is given by (where π_i is the fraction of each "type" i):

$$\overline{D}_{jk}(x) = \sum_{i=1}^{I} \pi_i \frac{s_{ik}(x)}{1 - s_{ij}(x)} \cdot \frac{s_{ij}(x)}{s_j(x)}.$$

This paper: Linear Algebra Notation

Individual i's share for each choice given by $\mathbf{s_i} = [s_{i0}, s_{i1}, \ldots, s_{iJ}]$ and aggregate shares by $\sum_{i=1}^{I} \pi_i \cdot \mathbf{s_i} = \mathbf{s}$. The individual diversion ratio is given by $\mathbf{D}_i = \mathbf{s_i} \cdot \left[\frac{1}{(1-\mathbf{s_i})}\right]^T$. We write the $(J+1) \times (J+1)$ matrix of second-choice diversion as:

$$\begin{aligned}
\mathbf{D} &= \left(\sum_{i=1}^{I} \pi_i \cdot \mathbf{s_i} \cdot \left[\frac{1}{(1-\mathbf{s_i})}\right]^T \cdot \text{diag}(\mathbf{s_i})\right) \cdot \text{diag}(\mathbf{s})^{-1} \\
&= \left(\sum_{i=1}^{I} \pi_i \cdot \mathbf{s_i} \cdot \mathbf{s_i}^T \cdot \text{diag}(1-\mathbf{s_i})^{-1}\right) \cdot \text{diag}(\mathbf{s})^{-1}
\end{aligned}$$

This paper: What is the point?

Under relatively general conditions, second-choice diversion can be written as:

$$\mathbf{D} = \left(\sum_{i=1}^{I} \pi_i \cdot \mathbf{s_i} \cdot \mathbf{s_i}^T \cdot \text{diag}(1 - \mathbf{s_i})^{-1} \right) \cdot \text{diag}(\mathbf{s})^{-1}$$

- The (unrestricted) matrix of diversion ratios $\mathbf{D}$ is $(J+1) \times (J+1)$.
- Each individual diversion ratio is of rank one since it is the outer product of $\mathbf{s_i}$ with itself (and some diagonal "weights").
- Logit restricts $\mathbf{D}$ to be of rank one. Nested logit of rank $\leq G$ (the number of non-singleton nests). Mixed logit to rank$(\mathbf{D}) \leq I$ (but bound is likely uninformative).

Setting

- Assume that we observe aggregate market shares $\mathcal{S}_j$ and some subset of the diversion matrix $\mathcal{D}_{jk}$ for $(j,k) \in$ OBS.
- Goal: Can we obtain an estimate for the remainder of the matrix $\mathcal{D}$?
  - Related to CS literature on matrix completion methods.
- Because of the structure for $\mathbf{D}$ we know how to look for low-rank approximations:

$$\mathbf{D} = \left(\sum_{i=1}^{I} \pi_i \cdot \mathbf{s_i} \cdot \mathbf{s_i}^T \cdot \text{diag}(1-\mathbf{s_i})^{-1}\right) \cdot \text{diag}(\mathbf{s})^{-1}$$

  - Use the outer product of vectors $\mathbf{s_i} \cdot \mathbf{s_i}^T$ and increase "types" I.

Our Semiparametric Problem

$$\min_{s_{ij},\pi_i,\overline{q}_0} \sum_{(k,j)\in \mathsf{OBS}} \left(\mathcal{D}_{kj} - D_{kj}\right)^2 + \lambda_1 \sum_j (\mathcal{S}_j - s_j)^2 + \lambda_2 \left\|w_i\right\|^2$$

$$\text{subject to} \quad s_j = \sum_{i=1}^{I} \pi_i \cdot s_{ij}$$

$$D_{kj} = \sum_{i=1}^{I} \pi_i \cdot \frac{s_{ij}}{1 - s_{ik}} \cdot \frac{s_{ik}}{s_k}$$

$$\mathcal{S}_j = \frac{q_j}{\overline{q}_0 + \sum_{j\in\mathcal{J}_t} q_j}$$

$$0 \le s_{ij}, \pi_i, s_j, D_{kj} \le 1, \quad \sum_{i=1}^{I} \pi_i = 1, \quad \sum_j s_{ij} = 1$$

Estimate with cross-validation to select types I.

Discussion

- Goal: a good predictive model for withheld elements of $\mathcal{D}$.
- We are worried about overfitting so we use cross validation (withholding rows of $\mathcal{D}$) to select number of types I.
  - Otherwise we will always prefer the more complicated model
  - Compare models based on out of sample fit (LL, AIC, MSE, MAE).
- Model may or may not be sparse $s_{ij} = 0$ for some (i, j)
  - Could be that consumer i doesn't consider j.
  - Or consequence that $s_{ij} \geq 0$ and $\sum_j s_{ij} = 1$ amounts to an L_1 penalty $\sum_j |s_{ij}| \leq 1$.s
- Model is a semiparametric-logit since V_{ij} is unrestricted:

$$u_{ij} = V_{ij} + \varepsilon_{ij}, \quad s_{ij} = \frac{e^{V_{ij}}}{1 + \sum_k e^{V_{ik}}}$$

Comparison: Fox, Kim, Ryan, Bajari (QE 2011)

$$\min_{\pi_i} \sum_j \left(\mathcal{S}_j - \sum_i \pi_i \cdot \hat{s}_{ij}(\widehat{\beta}_i) \right)^2 \quad \text{subject to} \quad \hat{s}_{ij}(\widehat{\beta}_i) = \frac{e^{\widehat{\beta}_i x_j}}{1 + \sum_{j'} e^{\widehat{\beta}_i x'_j}}$$

$$0 \leq \pi_i \leq 1, \quad \sum_i \pi_i = 1$$

- Draw $\beta_i \sim G(\beta_i)$ from a prior distribution.
- Solved in characteristic space with a semi-parametric form for $F(\beta_i)$.
- Often produces very sparse models $\pi_i = 0$ (for all but 50 of 1000 simulated consumers).
- Hard to incorporate fixed parameters.

Comparison: Raval et al (2017, 2020)

- Cut data into bins (zip, income, age, gender)
- Observe shares (hospital demand) within each bin $s_{g(i),j}$
- A separate plain logit for each bin with only ξ_j as the common parameter.
- Use second choices from hospital closures (natural disasters) to compare models.

$$s_{g(i),j} = \frac{e^{\beta_g x_j + \xi_j}}{1 + \sum_{j'} e^{\beta_g x_{j'} + \xi_{j'}}}, \quad D_{kj,i} = \frac{s_{g(i),j}}{1 - s_{g(i),k}}$$

Latent Class Logit (Greene and Hensher 2003)

Most similar to what we're doing here.

- Estimate separate β_i for each class.
- Estimate proportion of each class π_i.
- Estimating finite mixtures is tricky and usually requires EM.

$$s_j(\pi, \beta) = \sum_{i=1}^{I} \pi_i \cdot \left(\frac{e^{\beta_i x_{ij} + \xi_j}}{1 + \sum_k e^{\beta_i x_{ik} + \xi_k}} \right)$$

Data

Description of Data

- Same data as Conlon and Mortimer AUD (2021).
- Not the same as Conlon and Mortimer AEJ (2013).
- 66 Vending Machines in white-collar office buildings in downtown Chicago
- About 35-40 snack products in each
- 10 experiments (2.5-3.5 weeks long) where either one or two top selling products from a single category were removed
  - Snickers, M&M Peanut, Animal Crackers, Famous Amos Cookies, Doritos Nacho, Cheetos
  - Snickers + M&M Peanut and Doritos Nacho+ Cheetos

Calculating Treatment Effects

Develop a matching estimator for Δq_k and Δq_j based on the following

1. A control week is valid IFF it is from the same machine.
2. A control week is valid IFF total sales do not increase, and do not decrease more than expected sales of removed product. $Q_s \in [Q_t, Q_t + E[q_j|Q_t]]$
3. D_{jk} is restricted to Unit Simplex. (Dirichlet Prior).

Restrict things to the simplex and do some Empirical Bayes Shrinkage for noisy estimates (infrequently stocked substitutes).

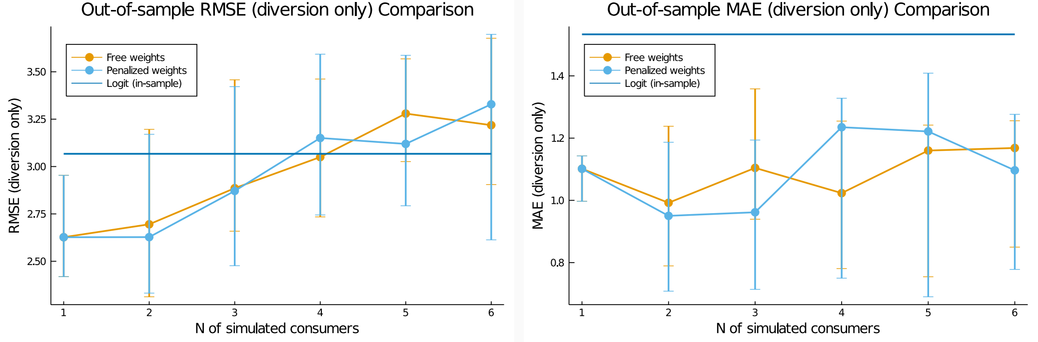
Experimental Diversion Estimates: Snickers Removal

Product	Raw	Nonparam
Outside Good	45.06	24.42
Twix Caramel	43.53	20.25
M&M Peanut 1.74 oz	31.96	16.44
Rold Gold (Con)	12.74	5.93
M&M Milk Chocolate	19.72	5.3
Butterfinger	13.63	4.13
Planters (Con)	9.32	3.96
Choc Chip Famous Amos	5.96	2.87
Choc Herhsey (Con)	21.27	2.57
Sun Chip LSS	4.89	2.23
Choc SandFamous Amos	7.62	1.75

Use this as data $\mathcal{S}$ and $\mathcal{D}$ for our estimates.

Results

Cross Validation: Model Selection



Mostly outperform plain logit. Seems to select $I = 3$ or $I = 4$.

Top Substitutes: Snickers

Product	Nonparam	Logit	CMS(I=3)	CMS(I=4)
Outside Good	24.42	26.27	29.53	26.21
Twix Caramel	20.25	2.38	11.49	16.38
M&M Peanut 1.74 oz	16.44	3.33	9.22	6.55
Rold Gold (Con)	5.93	2.88	2.65	4.32
M&M Milk Chocolate	5.3	1.86	6.95	5.39
Butterfinger	4.13	1.14	4.03	3.59
Planters (Con)	3.96	2.14	5.14	3.5
Choc Chip Famous Amos	2.87	1.63	0.49	0.42
Choc Herhsey (Con)	2.57	1.26	2.05	2.06
Sun Chip LSS	2.23	1.87	1.1	0.73
Choc SandFamous Amos	1.75	0.95	1.59	1.3

Top Substitutes: Doritos and Cheetos

Product	Nonparam	Logit	CMS(I=3)	CMS(I=4)
Outside Good	35.57	26.55	32.27	31.64
Baked (Con)	8.85	2.08	3.73	3.85
FritoLay (Con)	7.74	1.55	4.03	4.36
Snickers 2.07oz	7.52	3.7	1.64	3.05
Ruffles (Con)	6.77	2.54	4.21	4.62
Frito LSS	6.74	2.12	3.98	4.38
Dorito Blazin Buffalo Ranch LSS	3.07	1.52	1.55	1.57
Cheez-It Original SS	2.7	1.86	3.28	3.44
Rold Gold (Con)	2.66	2.92	1.38	1.71
Rice Krispies Treats 1.7oz	2.59	0.89	3.07	3.12
KarNuts (Con)	2.49	1.7	1.12	0.95

Model:		N = 1	N = 2		N = 3			N = 4			
Weight on individual:		100.00%	50.77%	49.23%	34.09%	33.66%	32.25%	25.81%	25.09%	24.76%	24.34%
productname	logit_sh	i = 1	i = 1	i = 2	i = 1	i = 2	i = 3	i = 1	i = 2	i = 3	i = 4
Cheez-It Original SS	1.77	1.47	0.75	3.09	4.7	0	1.47	4.64	0.49	0	1.35
Frito LSS	2.03	1.41	0.98	2.33	6.36	0	0	6.24	0	0.58	0
Dorito Nacho LSS	2.05	0.98	0.93	0.51	0	0	0	0	0	0	0
Cheeto LSS	2.52	0.51	0	0	0.78	0.19	0.3	1.05	0.4	0.01	0.07
Smartfood LSS	1.51	0.55	0.53	0.49	0.87	0.33	0.6	0.84	0.36	0.55	0.49
Sun Chip LSS	1.81	4.76	0	18.6	0	0	21.81	5.82	0	0	20.59
Cheetos Flaming Hot LSS	0.96	0.55	0.35	0.93	2.44	0	0.07	2.44	0	0.11	0.06
Lays Potato Chips 1oz SS	1.45	0.53	0.4	0.75	2.09	0	0.04	1.96	0	0.76	0
Dorito Blazin Buffalo Ranch I	1.45	1.47	0	5.49	0	0	6.8	1.83	0	0	6.36
Baked (Con)	1.99	2.09	2.39	1.09	5.64	1.28	0	4.6	2.41	0	0
FritoLay (Con)	1.49	1.71	1.35	2.42	6.25	0.03	0.49	6.14	0.11	1.56	0.42
Jays (Con)	1.48	0.17	0.16	0.13	0.33	0.08	0.17	0.3	0	0.72	0.1
Popcorn (Con)	2.06	0.51	0.57	0.22	0	0.63	0.69	0.13	0.42	0.93	0.58
Rold Gold (Con)	0.94	1.85	2.35	0.4	1.36	2.48	0.67	2.36	0	8.38	0.25
Salty Other (Con)	2.78	0.22	0.17	0.26	0.43	0.09	0.25	0.38	0	0.92	0.17
Snyders (Con)	2.21	0.7	0.34	1.52	2.63	0	0.34	2.44	0	1.26	0.28
Ruffles (Con)	0.98	1.88	0.61	4.92	5.62	0	3.06	6.39	0	0	2.97
Butterfinger	1.1	2.73	3.33	1.01	1.18	3.73	1.8	1.02	3.6	3.19	1.5
M&M Peanut 1.74 oz	3.21	4.78	7.09	0	2	8.74	0.02	0	10.22	2.25	0
M&M Milk Chocolate	1.8	3.62	5.32	0	0.39	6.58	0.02	0	6.92	3.42	0
Reeses Peanut Butter Cups	1.68	1.62	3.51	0	0	4.74	0	0	3.56	4.08	0
Snickers 2.07oz	3.53	6.13	8.68	0.01	0	9.99	0.55	0.77	9.54	9.23	0.49
Twix Caramel	2.29	5.19	8.1	0	0	10.89	0	0	0	32.51	0
Raisinets	1.47	1.46	2.19	0	0	2.74	0	0	3.45	0	0
Choc Herhsey (Con)	1.22	2.87	1.45	6.28	1.34	1.6	7.09	3.35	1.25	2.21	6.48
Choc Mars (Con)	2.13	1.03	1.57	0	0	2.05	0	0.36	0	6.49	0
Skittles Original	1.03	0.12	0.13	0	0.23	0.06	0.1	0.22	0	0.48	0.04
Twizzlers	1.66	1.15	1.14	1.06	2.09	0.91	0.6	1.26	1.76	0	0.66
Nonchoc Other (Con)	1.06	0.41	0.59	0	0.16	0.69	0.02	0.26	0	2.27	0
Zoo Animal Cracker Austin	1.9	0.29	0.37	0.05	0.02	0.62	0.14	0	0.53	0.25	0.17
Choc Chip Famous Amos	1.58	1.57	0	3.36	0	0	9.63	0	0.03	0	16.97
Rasbry Knotts	0.68	1.1	0.4	2.76	0.35	0.44	3.3	1.15	0.63	0	3.08
Choc SandFamous Amos	0.91	1.35	1.1	1.85	0	1.36	2.95	0	1.39	0.99	2.94
Grandmas Choc Chip	1.15	0.83	0.39	1.89	0.17	0.35	2.58	0.67	0.55	0	2.42
Ruger Wafer (Con)	1.6	0.54	0.62	0.25	1.05	0.47	0.08	1.18	0	2.18	0
Nabisco (Con)	1.23	1.44	1.13	2.07	0	1.29	3.13	0.06	1.97	0	2.85
Rice Krispies Treats 1.7oz	0.85	2.24	2.61	1.23	4.29	2.05	0.32	3.17	3.44	0	0.24
Pop-Tarts (Con)	2.42	0.27	0.37	0	0.21	0.38	0.1	0.24	0	1.47	0.02
Cherry Fruit Snacks	0.52	0.09	0.12	0	0.15	0.07	0.05	0.16	0	0.33	0
Ritz Bits Chs Vend	0.51	0.15	0.21	0	0.09	0.19	0.11	0.17	0	0.62	0.04
Farleys Mixed Fruit Snacks	0.99	0.58	0.45	0.8	2.27	0.03	0	2.01	0.28	0.14	0
KarNuts (Con)	1.65	1.25	1.77	0	1.37	1.73	0	0.41	2.57	0	0
Planters (Con)	1.63	4.81	3.54	7.77	0	4.37	10.56	0.88	6.09	0	9.93
Cliff (Con)	3.91	1.03	1.22	0.41	0	1.35	1.25	0	1.84	0	1.04
Nature Valley (Con)	2.13	1.42	1.29	1.58	5	0.37	0	3.93	1.4	0	0
Outside Good	25.34	28.58	29.43	24.49	38.13	27.09	18.85	31.18	34.81	12.12	17.44

	Cheez-It Original SS	Frito LSS	Dorito Nacho LSS	Cheeto LSS	Smartfood LSS	Sun Chip LSS	Cheetos Flaming Hot LSS	Lays Potato Chips 1oz SS	Dorito Blazin Buffalo Ranch LSS	Baked (Con)	FritoLay (Con)	Jays (Con)	Popcorn (Con)	Rold Gold (Con)	Salty Other (Con)	Snyders (Con)	Ruffles (Con)	Butterfinger	M&M Peanut 1.74 oz	M&M Milk Chocolate	Reeses Peanut Butter Cups	Snickers 2.07oz	Twix Caramel	Raisinets	Choc Herhsey (Con)	Choc Mars (Con)	Skittles Original	Twizzlers	Nonchoc Other (Con)	Zoo Animal Cracker Austin	Choc Chip Famous Amos	Rasbry Knotts	Choc SandFamous Amos	Grandmas Choc Chip	Ruger Wafer (Con)	Nabisco (Con)	Rice Krispies Treats 1.7oz	Pop-Tarts (Con)	Cherry Fruit Snacks	Ritz Bits Chs Vend	Farleys Mixed Fruit Snacks	KarNuts (Con)	Planters (Con)	Cliff (Con)	Nature Valley (Con)	Outside Good
Cheez-It Original SS	0	5.02	1.37	3.28	2.81	1.88	4.73	4.74	1.58	4.07	4.75	3.15	0.76	1.67	3.14	4.46	3.8	1.24	0.9	0.27	0	0.07	0	0	1.77	0	3.13	3.06	0.91	0.4	1.63	1.64	1.02	1.52	3.2	1.06	3.25	1.65	2.91	1.54	4.75	2.12	1.15	0.7	4.61	3.89
Frito LSS	5.15	0	1.12	3.98	3.16	0	6.35	6.37	0	5.5	6.29	3.69	0	1.98	3.62	5.82	4.45	1.15	1.22	0.36	0	0	0	0	0.9	0	3.74	3.81	1.19	0.2	0	0.58	0	0.37	4.23	0	4.31	1.96	3.61	1.55	6.43	2.87	0	0	6.23	4.74
Dorito Nacho LSS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cheeto LSS	0.7	0.83	0.34	0	0.52	0.38	0.78	0.78	0.32	0.71	0.79	0.56	0.25	0.39	0.56	0.74	0.65	0.33	0.32	0.24	0.2	0.22	0.22	0.2	0.37	0.2	0.56	0.56	0.3	0.23	0.33	0.34	0.27	0.32	0.59	0.28	0.6	0.39	0.54	0.36	0.79	0.46	0.3	0.25	0.77	0.74
Smartfood LSS	0.85	0.93	0.53	0.74	0	0.77	0.89	0.89	0.65	0.82	0.91	0.73	0.47	0.55	0.73	0.87	0.82	0.51	0.46	0.39	0.35	0.38	0.37	0.34	0.63	0.34	0.72	0.71	0.44	0.4	0.67	0.62	0.53	0.6	0.71	0.54	0.72	0.54	0.69	0.54	0.89	0.58	0.58	0.47	0.88	0.96
Sun Chip LSS	5.06	0	8.07	4.98	7.06	0	0.56	0.41	23.4	0	1.5	6.25	11.24	3.14	6.84	2.38	7.64	5.76	0.04	0.05	0.02	1.1	0	0.01	16.36	0	5.42	3.51	0.45	3.68	24.13	18.04	15.17	18.44	1.06	15.75	0.99	2.97	3.63	5.82	0	0	17.04	10.4	0	5.79
Cheetos Flaming Hot LSS	1.99	2.61	0.46	1.54	1.24	0.09	0	2.45	0.07	2.11	2.42	1.44	0.04	0.77	1.41	2.24	1.73	0.46	0.47	0.14	0	0	0	0	0.4	0	1.45	1.48	0.46	0.09	0.08	0.28	0.05	0.2	1.63	0.05	1.66	0.76	1.4	0.61	2.47	1.1	0.05	0.03	2.39	1.84
Lays Potato Chips 1oz SS	1.7	2.23	0.38	1.32	1.05	0.05	2.09	0	0.05	1.81	2.07	1.22	0.02	0.65	1.2	1.92	1.48	0.39	0.4	0.12	0	0	0	0	0.33	0	1.24	1.26	0.39	0.07	0.05	0.22	0.03	0.16	1.39	0.03	1.42	0.65	1.19	0.52	2.11	0.94	0.03	0.02	2.05	1.57
Dorito Blazin Buffalo Ranch LSS	1.58	0	2.52	1.55	2.2	8.69	0.17	0.13	0	0	0.47	1.95	3.5	0.98	2.13	0.74	2.38	1.8	0.01	0.01	0	0.34	0	0	5.1	0	1.69	1.09	0.14	1.15	7.52	5.62	4.73	5.74	0.33	4.91	0.31	0.93	1.13	1.81	0	0	5.31	3.24	0	1.81
Baked (Con)	4.57	6.03	1.58	3.73	3.04	0	5.64	5.66	0	0	5.59	3.45	0.63	2.48	3.37	5.17	3.95	1.77	2.22	1.61	1.35	1.35	1.44	1.32	1.01	1.31	3.53	3.71	2.08	1.21	0	0.65	0.42	0.48	4.14	0.39	4.23	2.46	3.55	2.01	5.72	3.27	0.4	0.69	5.62	4.77
FritoLay (Con)	5.17	6.67	1.3	4.03	3.27	0.62	6.25	6.27	0.52	5.41	0	3.77	0.27	2.03	3.72	5.77	4.54	1.28	1.22	0.39	0.03	0.05	0.03	0.03	1.25	0.03	3.8	3.83	1.2	0.3	0.54	0.97	0.35	0.78	4.19	0.36	4.27	2.01	3.63	1.67	6.32	2.83	0.39	0.25	6.13	4.8
Jays (Con)	0.31	0.35	0.16	0.26	0.23	0.22	0.33	0.33	0.19	0.3	0.34	0	0.13	0.17	0.25	0.32	0.29	0.15	0.13	0.1	0.08	0.09	0.09	0.08	0.19	0.08	0.25	0.25	0.13	0.1	0.19	0.18	0.15	0.17	0.25	0.15	0.26	0.17	0.24	0.17	0.34	0.19	0.16	0.12	0.33	0.33
Popcorn (Con)	0.16	0	0.54	0.25	0.34	0.88	0.02	0.01	0.74	0.12	0.05	0.28	0	0.45	0.29	0.08	0.24	0.55	0.56	0.64	0.66	0.7	0.71	0.65	0.62	0.64	0.28	0.27	0.52	0.62	0.76	0.64	0.69	0.66	0.22	0.69	0.23	0.44	0.28	0.49	0.01	0.36	0.74	0.67	0.04	0.46
Rold Gold (Con)	1.26	1.45	1.61	1.38	1.36	0.86	1.38	1.38	0.72	1.64	1.4	1.32	1.56	0	1.28	1.32	1.19	1.87	2.46	2.58	2.6	2.65	2.78	2.55	1.11	2.53	1.38	1.56	2.25	2.15	0.74	0.95	1.28	0.94	1.67	1.24	1.73	1.89	1.55	1.73	1.41	2.02	1.31	1.65	1.51	2.29
Salty Other (Con)	0.4	0.46	0.21	0.34	0.31	0.32	0.43	0.43	0.27	0.39	0.44	0.33	0.17	0.22	0	0.42	0.39	0.2	0.16	0.12	0.1	0.11	0.1	0.09	0.26	0.09	0.33	0.32	0.16	0.13	0.27	0.25	0.2	0.25	0.32	0.21	0.33	0.22	0.31	0.21	0.43	0.24	0.22	0.17	0.42	0.42
Snyders (Con)	2.21	2.81	0.59	1.72	1.42	0.43	2.64	2.64	0.36	2.28	2.63	1.63	0.18	0.87	1.61	0	1.96	0.57	0.5	0.15	0	0.02	0	0	0.63	0	1.63	1.63	0.5	0.14	0.38	0.52	0.24	0.44	1.77	0.25	1.8	0.86	1.55	0.73	2.66	1.19	0.27	0.16	2.58	2.05
Ruffles (Con)	5.26	6	2.12	4.21	3.78	3.91	5.69	5.69	3.28	4.86	5.77	4.14	1.58	2.19	4.16	5.48	0	1.83	1.08	0.33	0	0.15	0	0	3.08	0	4.06	3.86	1.12	0.69	3.38	3.04	2.12	2.91	3.89	2.21	3.95	2.15	3.7	2.19	5.68	2.53	2.39	1.46	5.51	5
Butterfinger	1.37	1.26	2.57	1.72	1.86	2.3	1.22	1.22	1.93	1.71	1.31	1.71	2.76	2.73	1.68	1.28	1.46	0	3.55	3.83	3.92	4.03	4.19	3.84	2.14	3.81	1.76	1.95	3.24	3.35	1.99	2.01	2.48	2.03	1.97	2.44	2.05	2.69	1.97	2.6	1.24	2.64	2.58	2.86	1.41	3.01
M&M Peanut 1.74 oz	1.62	2.13	4.32	2.58	2.63	0.03	1.99	2	0.02	3.35	2.01	2.35	4.3	5.56	2.19	1.83	1.4	5.46	0	8.92	9.17	9.22	9.81	8.98	1.76	8.92	2.62	3.43	7.35	7.1	0.03	1.16	2.9	1.17	3.9	2.68	4.08	5.5	3.47	4.78	2.12	5.84	2.77	4.7	2.56	5.37
M&M Milk Chocolate	0.32	0.42	3.06	1.25	1.42	0.02	0.39	0.39	0.02	1.56	0.42	1.13	3.24	3.84	1.02	0.36	0.28	3.91	5.92	0	6.91	6.95	7.39	6.76	1.17	6.72	1.32	1.92	5.33	5.31	0.02	0.78	2.18	0.82	2.2	2.02	2.32	3.8	1.98	3.33	0.47	3.9	2.09	3.54	0.84	3.22
Reeses Peanut Butter Cups	0	0	2.15	0.72	0.88	0	0	0	0	0.88	0.02	0.65	2.33	2.68	0.57	0	0	2.76	4.21	4.78	0	5	5.32	4.87	0.79	4.84	0.78	1.21	3.78	3.82	0	0.53	1.57	0.57	1.4	1.45	1.48	2.65	1.27	2.33	0.06	2.68	1.5	2.55	0.33	2.1
Snickers 2.07oz	0.13	0	4.73	1.64	2.04	0.71	0.01	0.01	0.59	1.85	0.08	1.51	5.19	5.72	1.37	0.06	0.19	5.97	8.88	10.07	10.48	0	11.21	10.26	2.08	10.2	1.79	2.64	7.98	8.13	0.61	1.56	3.68	1.65	2.96	3.44	3.14	5.5	2.76	5.05	0.12	5.65	3.58	5.62	0.69	4.58
Twix Caramel	0	0	4.94	1.65	2.03	0	0	0	0	2.02	0.05	1.48	5.35	6.15	1.3	0	0	6.35	9.68	10.97	11.42	11.49	0	11.19	1.82	11.12	1.8	2.78	8.69	8.76	0	1.2	3.59	1.29	3.2	3.32	3.4	6.08	2.91	5.34	0.13	6.16	3.43	5.84	0.75	4.83
Raisinets	0	0	1.24	0.42	0.51	0	0	0	0	0.51	0.01	0.37	1.35	1.55	0.33	0	0	1.6	2.43	2.76	2.87	2.89	3.07	0	0.46	2.8	0.45	0.7	2.18	2.2	0	0.3	0.9	0.33	0.81	0.83	0.86	1.53	0.73	1.34	0.03	1.55	0.86	1.47	0.19	1.22
Choc Herhsey (Con)	2.73	1.43	3.59	2.7	3.26	9.07	1.52	1.48	7.61	1.46	1.82	3.03	4.44	2.34	3.18	2	3.42	3.05	1.69	1.7	1.68	2.05	1.8	1.65	0	1.63	2.81	2.35	1.67	2.53	7.85	6.16	5.46	6.26	1.71	5.61	1.73	2.27	2.37	3	1.37	1.51	6.04	4.24	1.42	3.59
Choc Mars (Con)	0	0	0.93	0.31	0.38	0	0	0	0	0.38	0.01	0.28	1	1.16	0.25	0	0	1.19	1.82	2.06	2.15	2.16	2.3	2.1	0.34	0	0.34	0.52	1.63	1.65	0	0.23	0.67	0.24	0.6	0.62	0.64	1.14	0.55	1	0.02	1.16	0.65	1.1	0.14	0.91
Skittles Original	0.21	0.24	0.11	0.17	0.16	0.13	0.23	0.23	0.11	0.21	0.23	0.17	0.08	0.12	0.17	0.22	0.19	0.11	0.1	0.08	0.07	0.07	0.07	0.07	0.12	0.07	0	0.17	0.1	0.08	0.11	0.11	0.09	0.11	0.18	0.09	0.18	0.12	0.16	0.11	0.23	0.14	0.1	0.08	0.23	0.22
Twizzlers	1.83	2.23	1.01	1.58	1.4	0.77	2.1	2.11	0.65	1.98	2.11	1.51	0.76	1.25	1.49	1.98	1.67	1.07	1.21	1.04	0.96	0.99	1.02	0.94	0.9	0.93	1.53	0	1.13	0.9	0.67	0.79	0.72	0.74	1.69	0.71	1.73	1.24	1.53	1.12	2.13	1.46	0.76	0.78	2.11	2.12
Nonchoc Other (Con)	0.13	0.17	0.35	0.21	0.21	0.02	0.16	0.16	0.02	0.27	0.16	0.19	0.35	0.44	0.18	0.15	0.12	0.44	0.64	0.71	0.72	0.73	0.78	0.71	0.15	0.71	0.21	0.28	0	0.56	0.02	0.11	0.24	0.11	0.31	0.22	0.33	0.44	0.28	0.38	0.17	0.46	0.23	0.38	0.21	0.43
Zoo Animal Cracker Austin	0.05	0.03	0.33	0.14	0.17	0.17	0.03	0.03	0.15	0.14	0.04	0.14	0.37	0.38	0.13	0.04	0.06	0.4	0.55	0.62	0.65	0.66	0.69	0.64	0.21	0.63	0.15	0.19	0.5	0	0.15	0.18	0.3	0.19	0.2	0.29	0.22	0.37	0.2	0.35	0.03	0.36	0.3	0.4	0.07	0.33
Choc Chip Famous Amos	2.23	0	3.57	2.2	3.12	12.32	0.25	0.18	10.34	0	0.66	2.76	4.97	1.39	3.02	1.05	3.38	2.55	0.02	0.02	0.01	0.49	0	0.01	7.22	0	2.4	1.55	0.2	1.63	0	7.97	6.7	8.14	0.47	6.96	0.44	1.31	1.6	2.57	0	0	7.52	4.59	0	2.56
Rasbry Knotts	1.05	0.38	1.48	1.04	1.32	4.22	0.44	0.41	3.54	0.39	0.58	1.21	1.91	0.83	1.29	0.68	1.4	1.19	0.46	0.47	0.46	0.63	0.49	0.45	2.6	0.44	1.1	0.85	0.48	0.92	3.65	0	2.44	2.86	0.52	2.51	0.52	0.8	0.86	1.18	0.36	0.4	2.71	1.81	0.37	1.33
Choc SandFamous Amos	0.68	0	1.71	0.88	1.21	3.77	0.08	0.06	3.17	0.25	0.21	1.03	2.19	1.19	1.09	0.32	1.03	1.57	1.22	1.38	1.43	1.59	1.53	1.4	2.44	1.39	0.96	0.82	1.15	1.59	3.26	2.59	0	2.66	0.54	2.54	0.56	1.16	0.85	1.46	0.02	0.77	2.73	2.14	0.09	1.39
Grandmas Choc Chip	0.74	0.18	1.14	0.75	0.99	3.29	0.24	0.22	2.76	0.22	0.35	0.89	1.5	0.62	0.95	0.44	1.02	0.92	0.35	0.37	0.37	0.5	0.4	0.36	2.02	0.36	0.8	0.61	0.37	0.72	2.85	2.19	1.91	0	0.34	1.97	0.35	0.6	0.62	0.9	0.18	0.28	2.12	1.42	0.19	0.97
Ruger Wafer (Con)	0.87	1.12	0.43	0.75	0.64	0.1	1.05	1.06	0.09	1	1.05	0.7	0.27	0.61	0.68	0.97	0.76	0.49	0.62	0.54	0.5	0.5	0.53	0.49	0.29	0.48	0.72	0.76	0.58	0.43	0.09	0.21	0.21	0.19	0	0.2	0.86	0.6	0.74	0.51	1.07	0.74	0.21	0.29	1.06	1.02
Nabisco (Con)	0.73	0	1.74	0.91	1.25	4	0.08	0.06	3.36	0.24	0.22	1.07	2.25	1.18	1.14	0.34	1.1	1.58	1.15	1.3	1.35	1.52	1.45	1.32	2.56	1.31	0.99	0.83	1.09	1.56	3.46	2.73	2.6	2.8	0.53	0	0.54	1.15	0.86	1.47	0.02	0.73	2.85	2.18	0.09	1.4
Rice Krispies Treats 1.7oz	3.55	4.59	1.81	3.07	2.62	0.41	4.3	4.31	0.34	4.1	4.28	2.86	1.17	2.54	2.79	3.97	3.12	2.06	2.64	2.31	2.15	2.18	2.3	2.11	1.19	2.09	2.94	3.15	2.45	1.84	0.35	0.88	0.9	0.76	3.48	0.85	0	2.51	3.04	2.14	4.37	3.09	0.89	1.25	4.35	4.19
Pop-Tarts (Con)	0.19	0.22	0.24	0.21	0.2	0.12	0.21	0.21	0.1	0.25	0.21	0.2	0.23	0.29	0.19	0.2	0.18	0.28	0.37	0.39	0.4	0.4	0.42	0.39	0.16	0.38	0.21	0.24	0.34	0.33	0.11	0.14	0.19	0.14	0.25	0.18	0.26	0	0.23	0.26	0.21	0.31	0.19	0.25	0.23	0.35
Cherry Fruit Snacks	0.13	0.16	0.08	0.12	0.1	0.06	0.15	0.15	0.05	0.14	0.15	0.11	0.06	0.09	0.11	0.14	0.12	0.08	0.09	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.11	0.12	0.09	0.07	0.05	0.06	0.06	0.06	0.12	0.06	0.13	0.09	0	0.08	0.15	0.11	0.06	0.06	0.15	0.16
Ritz Bits Chs Vend	0.1	0.1	0.14	0.11	0.12	0.14	0.1	0.1	0.12	0.12	0.1	0.11	0.15	0.15	0.11	0.1	0.11	0.16	0.19	0.2	0.2	0.2	0.21	0.19	0.13	0.19	0.11	0.12	0.17	0.17	0.12	0.12	0.14	0.12	0.12	0.14	0.13	0.15	0.12	0	0.1	0.15	0.14	0.15	0.1	0.18
Farleys Mixed Fruit Snacks	1.84	2.42	0.41	1.42	1.13	0	2.27	2.27	0	1.97	2.24	1.32	0.01	0.72	1.3	2.08	1.59	0.43	0.46	0.16	0.03	0.03	0.03	0.03	0.32	0.03	1.34	1.37	0.45	0.09	0	0.21	0.01	0.14	1.52	0.01	1.55	0.72	1.29	0.57	0	1.04	0.01	0.01	2.23	1.7
KarNuts (Con)	1.11	1.46	1.03	1.12	1	0	1.37	1.37	0	1.5	1.36	1.03	0.85	1.04	0.99	1.25	0.96	1.26	1.8	1.82	1.81	1.83	1.94	1.78	0.48	1.77	1.09	1.25	1.64	1.54	0	0.31	0.57	0.29	1.42	0.53	1.47	1.39	1.24	1.18	1.4	0	0.55	0.93	1.46	1.79
Planters (Con)	2.45	0	5.89	3.07	4.23	13.51	0.27	0.2	11.33	0.81	0.74	3.62	7.59	3.99	3.84	1.15	3.7	5.34	3.9	4.42	4.59	5.14	4.9	4.49	8.65	4.46	3.35	2.81	3.7	5.29	11.69	9.22	8.78	9.45	1.8	8.96	1.85	3.88	2.92	4.96	0.05	2.47	0	7.38	0.3	4.74
Cliff (Con)	0.29	0	1.07	0.49	0.66	1.6	0.03	0.02	1.34	0.25	0.09	0.54	1.31	0.94	0.55	0.14	0.44	1.12	1.2	1.36	1.42	1.49	1.51	1.39	1.16	1.38	0.53	0.55	1.1	1.3	1.39	1.18	1.32	1.22	0.46	1.31	0.48	0.92	0.57	1	0.02	0.76	1.4	0	0.09	0.93
Nature Valley (Con)	4.05	5.34	1.05	3.18	2.55	0	4.99	5.01	0	4.4	4.95	2.95	0.18	1.76	2.89	4.58	3.5	1.13	1.29	0.66	0.39	0.4	0.42	0.39	0.77	0.38	3	3.09	1.23	0.46	0	0.49	0.12	0.34	3.44	0.11	3.5	1.75	2.94	1.4	5.06	2.46	0.12	0.2	0	3.89
Outside Good	35.24	40.72	25.99	32.28	30.09	24.1	38.57	38.56	20.22	38.03	39.13	31.21	23.02	29.85	30.89	36.96	33.28	27.69	31.39	29.5	28.43	29.53	30.4	27.85	24.03	27.65	31.58	32.81	29.14	26.19	20.86	22.03	22.03	21.37	34.27	21.85	35.16	29.45	32.01	27.62	38.87	32.49	23.26	23.52	39.24	0

Conclusion

- Allowing for flexible unobserved types can give more accurate substitution patterns
  - Particularly true in capturing closeness of best substitutes not captured by product characteristics (e.g. Snickers and Peanut M&M's vs Snickers and Milky Way)
- Using observable substitution patterns (as measured from experiments or surveys) and "completing" the $(J+1) \times (J+1)$ matrix with a low-rank approximation looks promising.
- How much information on second choices is "enough" to get reasonable estimates of the rest?
- Other cases where we have incomplete substitution patterns?