Estimating Preferences and Substitution Patterns from Second-Choice Data Alone

Chris Conlon (NYU Stern & NBER) and
Julie Holland Mortimer (University of Virginia & NBER) and
Paul Sarkis (Boston College)

May 29, 2023

Review of Diversion Ratios

Diversion Ratios

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

- Raise the price of product j and count the number of consumers who leave
- The diversion ratio $D_{j \to k}$ is the fraction of leavers who switch to the substitute k.
- A higher value of $D_{j \to k}$ indicates closer substitutes.
- Useful because it arises in the multi-product Bertrand FOC:

$$\underbrace{p_j \left(1 + 1/\epsilon_{jj}(\mathbf{p})\right)}_{\text{Marginal Revenue}} = c_j + \sum_{k \in \mathcal{J}_f \setminus j} (p_k - c_k) \cdot D_{j \to k}(\mathbf{p}).$$

- $D_{j \to k} \equiv \frac{\partial q_k}{\partial p_j} / \left| \frac{\partial q_j}{\partial p_j} \right|$.
- Can also write as $D_{j \to k} \equiv \frac{\epsilon_{kj}}{|\epsilon_{jj}|} \cdot \frac{q_j}{q_k}$

General Advantages of Diversion

- Diversion allows for unit-free comparisons (shares sum to one).
  - While own-elasticities are unit-free, this is not true of cross-elasticities.
  - Is $\epsilon_{jk} = .01$ or $\epsilon_{jk} = .03$ a better substitute? We can't tell.
    - Need $\epsilon_{jk} \cdot s_k$ to know.
    - But $\epsilon_{jk} \cdot \frac{s_k}{p_j} = D_{j \to k}$.
    - The fraction of switchers choosing k allows comparisons.
  - If tempted to report cross elasticities, consider reporting diversion ratios instead.
- Data on diversion can provide helpful variation for demand estimation.
  - Petrin (2002), MicroBLP (2004), Grieco, Murry, Yurukoglu (2022)
- Diversion can be a helpful complement to merger simulation.

Advantages of Diversion in Merger Analysis (Farrell and Shapiro, 2010)

Diversion vs. concentration:

- Most goods and services are differentiated.
- Merger policy should aim to measure the substitutability of the differentiated offerings of competing firms.
- Concentration measures typically struggle to do this:
    - not all firms "in the market" produce products that are equally good substitutes
    - some firms "outside the market" may produce products that compete.
- If merging parties know they compete more closely than market-share analysis would predict, we'll have under-enforcement.

Diversion as a Treatment Effect (Conlon Mortimer RJE 2021)

Diversion Ratio = fraction of consumers who switch from purchasing a product j to purchasing a substitute k *(following an increase in the price of j)*

Treatment not purchasing product j

Outcome fraction of consumers who switch from $j \rightarrow k$.

Compliers consumers who would have purchased at z_j but do not purchase at z_j'.

This admits a Wald estimator:

$$D_{j \to k}(x) = \frac{\mathbb{E}[q_k | Z = z_j'] - \mathbb{E}[q_k | Z = z_j]}{\mathbb{E}[q_j | Z = z_j] - \mathbb{E}[q_j | Z = z_j']}$$

A LATE Theorem (Conlon Mortimer RJE 2021)

We also showed that most discrete-choice models yield the following representation:

$$D_{j\to k}^{z_j\to z_j'}(x) = \int_{z_j}^{z_j'} D_{j\to k,i}(x)\, w_i(z_j, z_j', x)\, dF_i \text{ with } w_i(z_j, z_j', x) = \frac{s_{ij}(z_j, x) - s_{ij}(z_j', x)}{s_j(z_j, x) - s_j(z_j', x)}$$

- Different interventions $z_j \to z_j'$ (prices, quality, characteristics, assortment) give different weights $w_i(z_j, z_j', x)$ and thus different local average diversion ratios.
- Individual Diversion Ratios $D_{j\to k,i}(x)$ don't vary with the intervention (determined only by how i ranks 2nd and 3rd choices).
- That paper establishes the decomposition above and derives some properties.

A Special Case: Second Choices and Mixed Logit

If the underlying model is (any) mixed logit then:

$$D_{j\to k,i} = \frac{s_{ik}}{1 - s_{ij}}$$

And if the intervention is to eliminate j from the choice set $\mathcal{J}$

$$w_{ij} = \frac{s_{ij}}{s_j}$$

So that (where π_i is weight on each type: Monte Carlo/Quadrature/etc.)

$$D_{j\to k} = \sum_{i=1}^{I} \pi_i \cdot \frac{s_{ik}}{1 - s_{ij}} \cdot \frac{s_{ij}}{s_j}$$

Which it turns out is very convenient.

Motivation

Motivation #1: Challenges of Parametric Models

Julie and I have been thinking about how consumers substitute across products for quite a while

- We've looked at parametric models from lots of studies
- We ran some experiments on vending machines machines
- Even the most complicated random coefficients models suffer from three main deficiencies:
  - Never quite enough substitution to best substitutes
  - Everything looks a bit too much like plain logit (substitution proportional to share)
  - Your substitution patterns are only as good as your characteristics → if you want extreme substitution patterns you need extreme characteristics.

Motivation #2: What is the right diversion ratio

Moreso in Europe than US...

- Agencies have data on substitution
  - Customer switching or win/loss of cell phone companies.
  - Second choice surveys (UK CMA does a lot of this)
  - "Course of business" diversion ratios (Farrel Shapiro 2010)
- These may not be the object we want to plug into the FOC as substitution patterns.
  - Often something like *ceteris paribus* response to small change in price.

Can we still use the information from the wrong experiment in a disciplined way?

Setup

Data

Consumers make discrete choices from set $\mathcal{J}$ and we observe market shares and select second-choice probabilities

$$\begin{aligned} \mathcal{S}_j &= \mathbb{P}(\text{chooses } j \in \mathcal{J}) \\ \mathcal{D}_{j\to k} &= \mathbb{P}(\text{chooses } k \in \mathcal{J} \setminus \{j\} \mid \text{chooses } j \in \mathcal{J}) \end{aligned}$$

We observe the set of (j,k) elements in $\mathcal{D}$ which we label P_Ω and its complement $P_{\overline{\Omega}}$.

Wrinkle: We observe data only from a single market.

Example: Cell Phone Merger

We consider a problem where we observe some aggregate shares $\mathcal{S} = [\mathcal{S}_1, \ldots, \mathcal{S}_J]$, and some elements of $\mathcal{D}^T$ a matrix of second-choice probabilities.

$$\mathcal{D}^T = \begin{array}{c c} \begin{array}{ccccc} \text{VZ} & \text{ATT} & \text{TMo} & \text{S} & \text{Other} \end{array} & \\ \begin{pmatrix} 0 & ? & 0.30 & 0.30 & ? \\ ? & 0 & 0.45 & 0.15 & 0 \\ ? & ? & 0 & 0.45 & ? \\ ? & ? & 0.20 & 0 & ? \\ ? & ? & 0.05 & 0.10 & 0 \end{pmatrix} & \begin{array}{c} \text{VZ} \\ \text{ATT} \\ \text{TMo} \\ \text{S} \\ \text{Other} \end{array} \end{array}, \quad \begin{bmatrix} 0.35 \\ 0.30 \\ 0.20 \\ 0.10 \\ 0.05 \end{bmatrix} = \mathcal{S}$$

Can we fill in the missing elements?

Can we estimate parameters and simulate the merger?

Definitions: First Choices

Utility is given by semi-parametric logit ε_{ij} is Type I extreme value $u_{ij} = V_{ij} + \varepsilon_{ij}$

Conditional choice probabilities (s_{ij}):

$$\mathbb{P}(u_{ij} > u_{ij'};\ \text{for all } j \neq j' \mid \mathbf{V_i}) = \frac{e^{V_{ij}}}{\sum_{j' \in \mathcal{J}} e^{V_{ij'}}} \equiv s_{ij}(\mathbf{V_i}).$$

Unconditional choice probabilities (s_j):

$$\mathbb{P}(u_{ij} > u_{ij'};\ \text{for all } j \neq j') = \int s_{ij}(\mathbf{V_i})\, f(\mathbf{V_i})\, \partial \mathbf{V_i} \approx \sum_{i=1}^{I} \pi_i\, s_{ij}(\mathbf{V_i}) \equiv s_j.$$

- $\mathbb{P}(\mathbf{V_i} = \mathbf{v_i}) = \pi_i$ so that $\pi_i \geq 0$ and $\sum_{i=1}^{I} \pi_i = 1$ (so that π constitutes a valid probability measure) for $f(\mathbf{V_i})$.
- Let $\mathbf{S}$ be a $\dim(\mathcal{J}) \times I$ matrix with column vectors $\mathbf{s_i}$ $\rightarrow$ We can write $\mathbf{s} = \mathbf{S}\,\pi$.

Definitions: Second Choices

For any (semiparametric) mixture of logits we can write the probability that individual i chooses k as their second choice given that j is their first choice as:

$$\begin{aligned} D_{j\to k} &\equiv \mathbb{P}(\text{ chooses } k \in \mathcal{J} \setminus \{j\} \mid \text{ chooses } j \in \mathcal{J}) \\ &= \sum_{i=1}^{I} \pi_i \cdot \frac{s_{ik}}{1-s_{ik}} \cdot \frac{s_{ij}}{s_j} \end{aligned}$$

It is convenient to interpret $D_{j\to k}$ as the (j,k)th entry in the second-choice matrix $\mathbf{D}(\mathbf{S},\pi)$.

Our Semiparametric Problem

$$\min_{(\mathbf{S},\pi)\geq 0} \left\|\mathcal{P}_{\Omega}(\mathcal{D} - \mathbf{D}(\mathbf{S},\pi))\right\|_{\ell_2} + \lambda \left\|\mathcal{S} - \mathbf{S}\,\pi\right\|_{\ell_2} \text{ with } \left\|\pi\right\|_{\ell_1} \leq 1, \quad \left\|\mathbf{s_i}\right\|_{\ell_1} \leq 1.$$

- Constraints: Choice probabilities s_{ij} sum to one, type weights π_i sum to one.
- Use cross validation to select $\#$ of types I and Lagrange multiplier λ.
- Not-convex but not very difficult either.
- ℓ_1 constraints lead to sparsity.
- Goal: estimate $\mathbf{s_i}$ (choice probabilities) and corresponding weights π_i (Finite Mixture)

Second Choice Matrix

- Individual i's share for each choice given by $\mathbf{s_i} = [s_{i0}, s_{i1}, \ldots, s_{iJ}]$.
- Aggregate shares by $\sum_{i=1}^{I} \pi_i \cdot \mathbf{s_i} = \mathbf{s}$.
- The matrix of individual diversion ratios 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 probabilities as:

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

Second Choice Matrix: Continued

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

$$\mathbf{D} = \text{diag}(\mathbf{s})^{-1} \cdot \left(\sum_{i=1}^{I} \pi_i \cdot \begin{bmatrix} | \\ \mathbf{s_i} \\ | \end{bmatrix} \cdot \begin{bmatrix} — & \frac{\mathbf{s_i}}{1-\mathbf{s_i}} & — \end{bmatrix} \right)$$

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

Second Stage

Recovering Price Sensitivity

Recovering substitution patterns is great but... What about price sensitivity?

- We need these to evaluate mergers, welfare, compute elasticities, etc.
- Need to impose some additional constraints, some options

$$V_{ij} - V_{i0} = \begin{cases} \ln \hat{s}_{ij} - \ln \hat{s}_{i0} & \text{if } \hat{s}_{ij} > 0, \\ \text{n.a.} & \text{if } \hat{s}_{ij} = 0. \end{cases}$$

- Can impose $V_{i0} = 0$ like everyone else.
- Remember these are at the individual level
- For now assume that $s_{ij} = 0$ is about consideration, otherwise have to deal with selection.

Recovering Price Sensitivity : Options

- Usual IV conditions

$$\min_{\beta_i^p, f_i(\cdot)} \left\| z_j'(\ln \hat{s}_{ij} - \ln \hat{s}_{i0} - f_i(x_j) - p_j \beta_i^p) \right\|_{\ell_2}.$$

- Matching observed elasticities (from another study, quasi-experimental estimate, etc.)

$$\min_{\beta_i^p < 0} \left\| \mathcal{E}_{jj} - \frac{p_j}{s_j} \sum_{i=1}^{I} \beta_i^p \cdot \widehat{\pi}_i \cdot \hat{s}_{ij} \cdot (1 - \hat{s}_{ij}) \right\|_{\ell_2}.$$

- Matching observed price-cost margins (antitrust agencies can subpoena..?)

$$\mathbf{c} = \mathbf{p} - \left(\mathcal{H} \odot \left(\sum_{i=1}^{I} \pi_i \cdot \Delta_i(\beta_i^p) \right) \right)^{-1} \mathbf{s},$$

$$\Delta_i(\beta_i^p) = \beta_i^p \left(-\mathbf{s_i} \, \mathbf{s_i}^T + \mathrm{diag}[\mathbf{s_i}] \right).$$

$$\min_{\beta_i^p < 0} \left\| \mathcal{C}_{jj} - c_j(\beta_i^p, \mathcal{H}, \hat{S}, \widehat{\pi}) \right\|_{\ell_2}.$$

Dicussion of Rank Restriction

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 second choices likely equally hopeless.
  - Maybe try LASSO or something to reduce cross terms?
- Plain logit places strong restrictions: $D_{j \to k} = \frac{s_k}{1-s_j}$.
- Nested logit $D_{j \to k} = \frac{s_{k|g}}{Z(\sigma, s_g) - s_{j|g}}$ (same nest) where σ is nesting parameter.

How do we fill in missing elements?

Mixed Logit: Explain substitution patterns using observed characteristics

- Typically assume independent normal RC
- Two products with similar x_1 and high substitution → larger σ_1.
- Two products with similar x_2 and low substitution → smaller σ_2.

McFadden and Train (2000) show a mixed logit $u_{ij} = \beta_i x_j + \varepsilon_{ij}$ is fully flexible

1. This depends on $f(\beta_i)$ heterogeneity being nonparametric
2. And a sufficient set of characteristics X to explain $\mathcal{D}$

Much work on (1), less attention on (2).

How do we fill in missing elements?

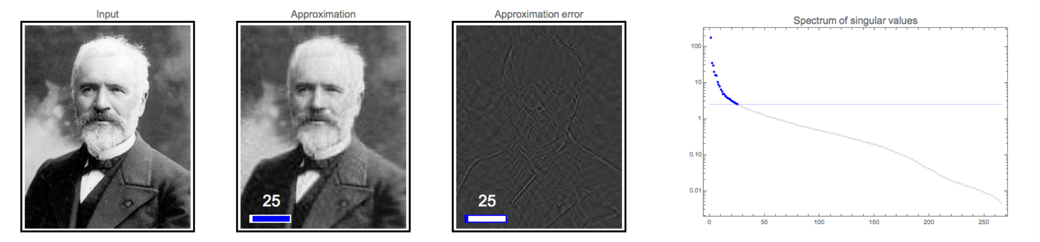
Our paper: Consider a low-rank approximation to $\mathcal{D}$

- Limit the rank of $\mathcal{D}$ directly in product space instead of controlling complexity with product characteristics and parametric restrictions on random coefficients.
- Allow for sparsity in individual shares and substitution patterns, with possibility of generating extreme patterns for top substitutes if necessary.

Works well in other domains (CS for image recovery/compression), and we show it has a sensible economic interpretation.

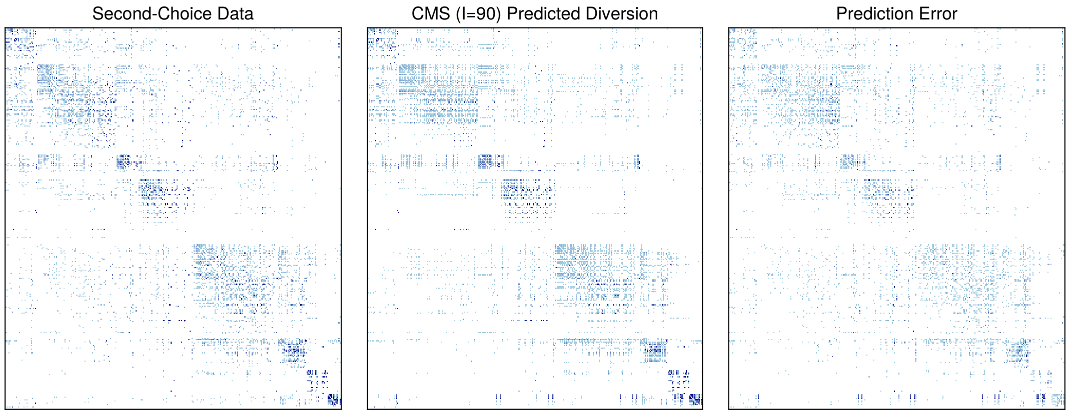
Low Rank Approximations: Image Compression

Image of Camille Jordan (1838-1922)



$$A \approx U_{266\times 25} \cdot \Sigma_{25\times 25} \cdot V_{25\times 266}$$

Completing the Matrix: $\mathcal{D}$ for Autos



When might we want to do this?

- We have access to aggregate market shares and some (but not all) second-choice data (microBLP (2004); Grieco, Murry, Yurukoglou (2022)).
- We are interested in estimating substitution patterns across all sets of products but have data on only a subset
  - shares of largest cellular phone providers, and number porting or switching data for merging parties only.
  - survey data on "If this Tesco were to close where would you shop" (as UK CMA asks).
  - win-loss data from merging parties only (Qiu, Sawada, Sheu (2022)) [not exactly]
- We lack sufficient variation in prices, other covariates, to estimate demand system.
- Product characteristics do not accurately capture substitution across products.

Comparisons

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

$$\min_{\pi_i \geq 0} \sum_j \left(\mathcal{S}_j - \sum_i \pi_i \cdot \hat{s}_{ij}(\widehat{\beta}_i) \right)^2 \quad \text{subject to} \quad \sum_i \pi_i = 1$$

$$\hat{s}_{ij}(\widehat{\beta}_i) = \frac{e^{\widehat{\beta}_i x_j}}{1 + \sum_{j'} e^{\widehat{\beta}_i x_j'}}$$

- 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).

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.

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

Comparison: Latent Class Logit (Greene and Hensher 2003)

Most similar to what we're doing.

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

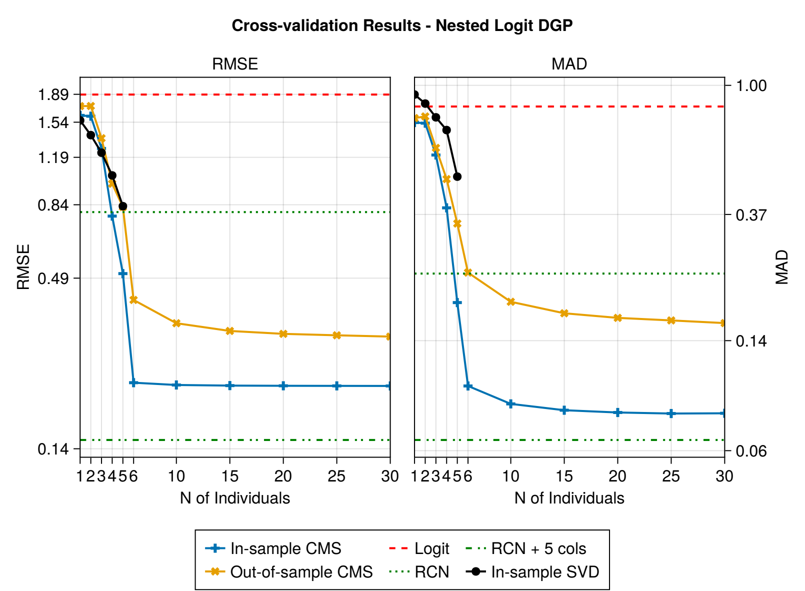
$$s_k(\pi, \beta) = \sum_{i=1}^{I} \pi_i \cdot \left(\frac{e^{\beta_i x_{ij} + \xi_j}}{1 + \sum_k e^{\beta_i x_{ik} + \xi_k}} \right)$$

Monte Carlo

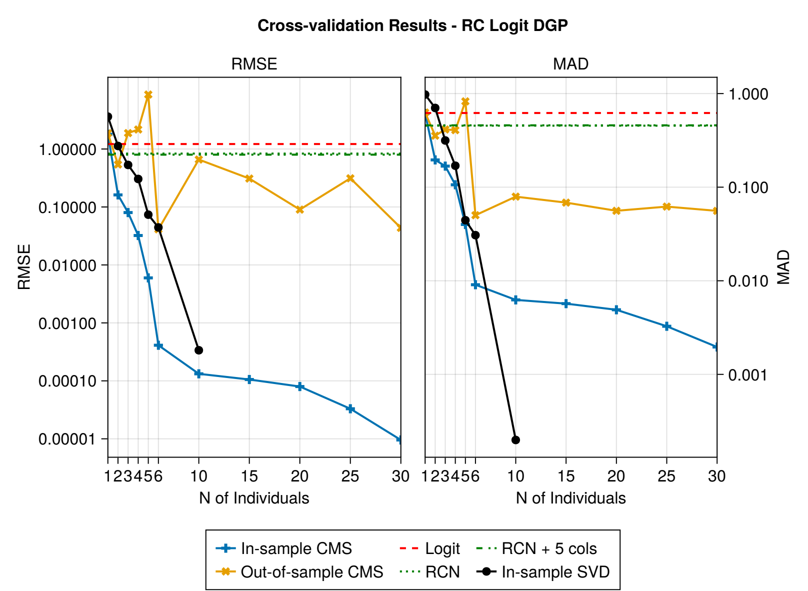
Generating Data

- Fit (i) nested logit, (ii) RC logit to data on vending machines from Conlon and Mortimer (JPE, 2021).
- Generate fake sales and diversion from those parameter estimates.
  - $J = 45$ products; $T = 250$ markets; with 30 randomly selected products in each. Market size $M = 1000$ per market. Nesting parameter is $\rho = 0.25$.
  - Categories: Salty Snacks, Chocolate, Non-Chocolate Candy, Cookies, Pastry, Other.
- Estimate a variety of misspecified parametric models: RC on nest dummies, RC on characteristics (Salt, Sugar, Nut Content), and our semiparametric estimator.
  - Include $m \ll J$ columns of $\mathcal{D}_{j\to k}$ as extra moments.
- Compare out-of-sample predicted Diversion Ratios.
  - MAD: Median$\left(|\mathcal{D}_{j\to k} - \hat{D}_{j\to k}|\right)$ for $(j,k) \in \{\text{Validation}\}$.
  - RMSE: $\sqrt{\frac{1}{n}\sum_{(j,k)\in\{\text{Validation}\}} |\mathcal{D}_{j\to k} - \hat{D}_{j\to k}|^2}$

Monte Carlo: DGP is Nested Logit



Monte Carlo: DGP is RC on chars

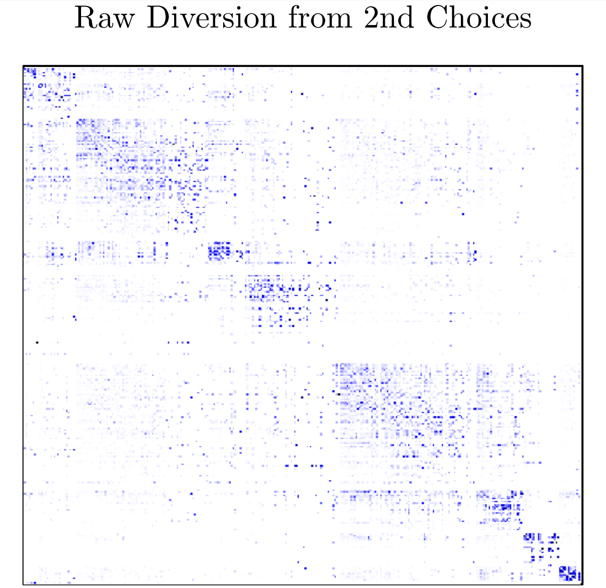


Application to Autos Data

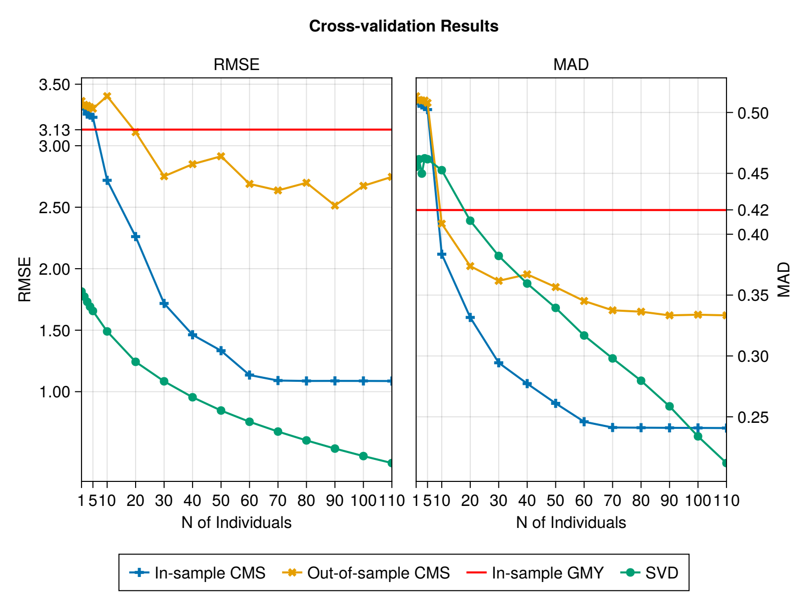
Description of Autos Data

- Subset of data from Grieco, Murry and Yurukoglu (2022).
- Focus on one year of sales from 2015
  - Aggregate sales observed at the model-year level from Ward's Automotive.
  - Second choices from MaritzCX survey (53,328 purchases)
  - In total, $J = 318$ products.
- Same Goal: Predict unobserved second-choice data without characteristics.

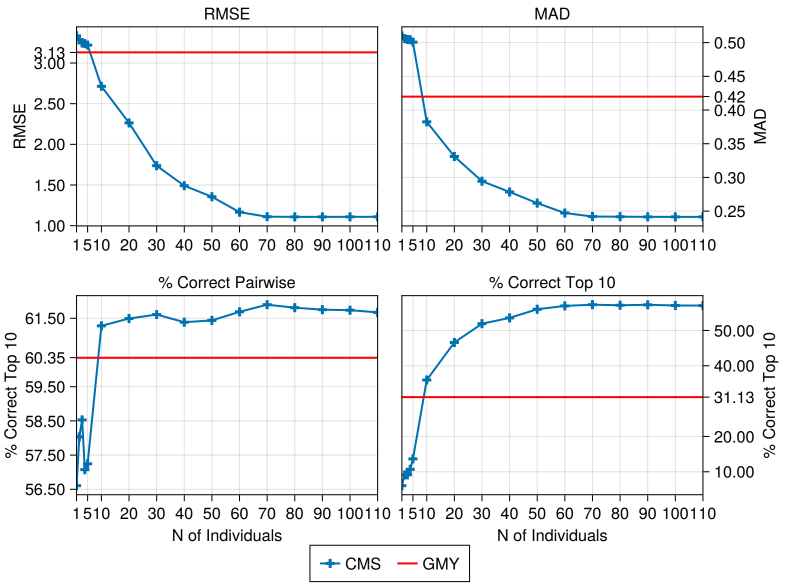
MaritzCX Survey data (318 Cars and Light Trucks)



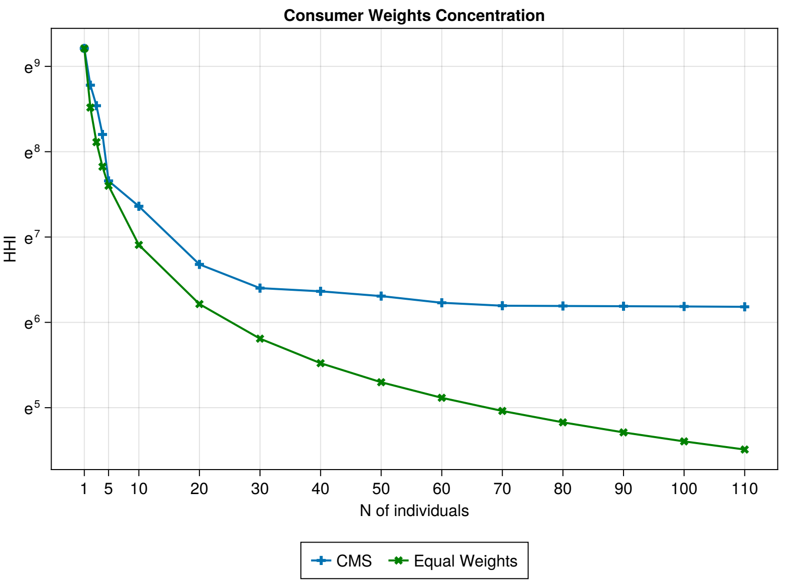
Cross Validation: Model Selection



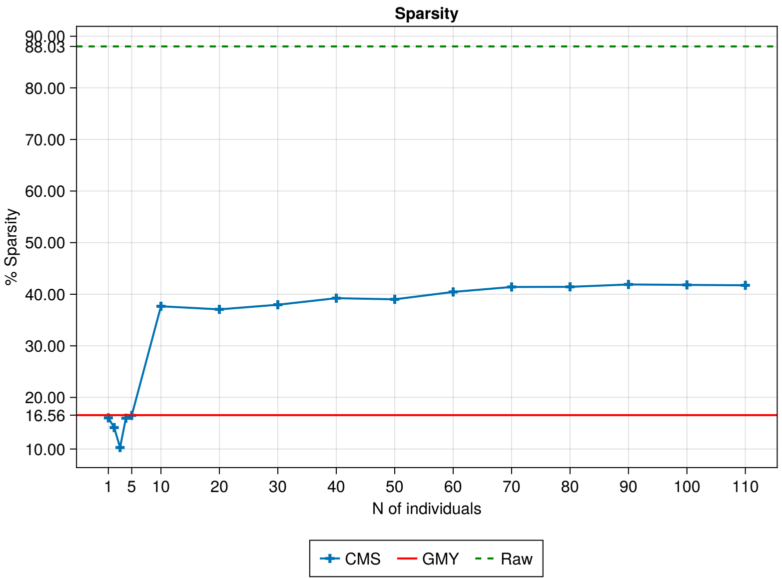
In-Sample Performance



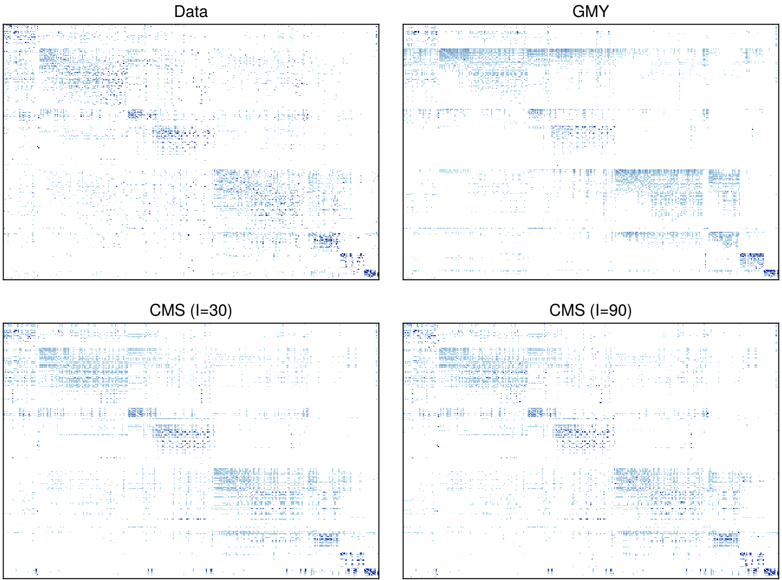
Analysis of Consumer Weights



Analysis of Sparsity



Comparison of Implied Diversion



Top Substitutes: Honda Accord

Model	Raw	Logit	CMS I=30	CMS I=90	GMY
Subaru Legacy	10.27	1.01	8.05	8.21	1.3
Toyota Camry	9.1	0.84	6.7	6.85	9.48
Acura Tlx	6.07	0.71	1.83	2.07	0.46
Honda Civic	5.97	0.91	2.9	2.75	3.89
Mazda Mazda6	5.68	0.52	4.77	4.87	1.32
Volkswagen Passat	4.01	0.74	4.34	4.41	1.22
Nissan Altima	3.52	0.6	3.87	3.89	7.22
Hyundai Sonata	3.52	0.68	5.61	5.68	5.09
Volkswagen Jetta	3.33	0.97	4.4	4.23	1.48
Mazda Mazda3	2.15	1.08	2.08	1.61	1.49
Toyota Corolla	1.96	0.71	2.46	2.32	4.66

Table 1: Top Substitutes: Honda Accord

Top Substitutes: Ford F-Series

Model	Raw	Logit	CMS I=30	CMS I=90	GMY
Ram Pickup	24.59	1.36	23.38	23.37	19.4
Gmc Sierra	20.29	1.28	21.0	21.02	17.27
Chevrolet Silverado	15.62	1.21	16.73	16.75	33.62
Toyota Tundra	12.98	0.76	12.69	12.69	2.29
Toyota Tacoma	6.31	1.13	3.6	3.62	2.83
Chevrolet Colorado	4.64	1.08	3.37	3.38	2.87
Gmc Canyon	2.3	0.62	1.71	1.73	1.02
Nissan Frontier	1.63	0.67	0.83	0.84	0.61
Jeep Wrangler	1.59	0.62	0.96	0.81	0.06
Nissan Titan	0.7	0.07	0.79	0.81	0.18
Ford Explorer	0.63	0.4	0.05	0.03	0.71

Table 2: Top Substitutes: Ford F-Series

Top Substitutes: Mercedes-Benz Sprinter Van

Model	Raw	Logit	CMS I=30	CMS I=90	GMY
Ford Transit Wagon	66.67	0.19	47.63	66.19	0.04
Ram Promaster	16.67	0.02	0.0	16.19	1.76
Ford Transit Connect	8.33	0.18	7.33	7.85	0.01
Nissan Nv	8.33	0.17	29.96	7.58	6.52
Mini Cooper	0.0	0.45	0.0	0.0	0.0
Volkswagen Beetle Ii Cabrio	0.0	0.25	0.0	0.0	0.0
Audi A5	0.0	0.28	0.0	0.0	0.02
Mazda Mx-5 Miata	0.0	0.19	0.0	0.0	0.0
Audi S5	0.0	0.15	0.0	0.0	0.0
Porsche Boxster	0.0	0.07	0.0	0.0	0.0
Volkswagen Eos	0.0	0.07	0.0	0.0	0.0

Table 3: Top Substitutes: Mercedes-Benz Sprinter Van

Vending Example

Product Removal Experiments

- Described in Conlon, Mortimer, Sarkis, Rodriguez-Valdenegro (2023)
- Used in Conlon Mortimer (JPE 2021) not (AEJM 2013)!
- Remove best sellers by category:
  - Chocolate: Snickers and M&M Peanut
  - Cookie: Animal Cracker and Famous Amos
  - Salty: Doritos and Cheetos
- 66 Vending machines in Downtown Chicago office buildings (around 10,000 treated individuals per arm)

Zoo Animal Crackers

Product	Shares	Nonparam	Logit	RCC	RCN	CMS(I=2)	CMS(I=3)
Outside Good	30.12	23.86	22.93	23.37	20.2	29.27	25.44
M&M Peanut 1.74 oz	4.14	9.72	3.34	3.74	2.43	5.94	7.31
Twix Caramel	2.42	7.81	2.44	2.91	1.79	6.72	9.0
Snickers 2.07oz	3.96	6.93	3.19	3.53	2.33	7.35	8.08
Planters (Con)	1.92	6.15	2.19	1.85	1.23	3.99	5.21
Choc Chip Famous Amos	2.05	5.99	1.66	1.87	8.23	0.31	4.74
Rold Gold (Con)	2.56	5.15	3.01	1.29	1.68	2.16	2.08
Choc Herhsey (Con)	0.22	3.68	1.31	1.63	1.0	2.07	2.42
Rice Krispies Treats 1.7oz	0.27	3.63	0.99	1.07	0.69	2.49	1.74
Baked (Con)	2.39	3.04	2.09	1.82	1.17	2.32	1.15
Popcorn (Con)	0.42	2.81	1.0	0.71	0.56	0.58	0.56

Snickers and M&M Peanut

Product	Shares	Nonparam	Logit	RCC	RCN	CMS(I=2)	CMS(I=3)
Outside Good	30.12	36.33	24.02	24.64	27.99	34.78	33.62
Twix Caramel	2.42	11.64	2.56	2.93	5.31	8.56	11.97
M&M Milk Chocolate	1.16	8.33	2.09	2.63	4.43	5.73	7.1
Choc Mars (Con)	1.11	6.79	1.76	2.11	3.68	1.71	2.22
Reeses Peanut Butter Cups	0.59	6.57	1.84	2.78	3.83	3.48	5.12
Butterfinger	0.5	5.22	1.22	1.71	2.49	3.75	4.2
Raisinets	1.6	3.28	1.67	2.12	3.48	2.38	2.92
Nonchoc Other (Con)	0.78	2.63	1.45	1.81	1.33	0.69	0.74
Choc Chip Famous Amos	2.05	2.48	1.74	1.79	1.23	0.0	0.21
Choc Herhsey (Con)	0.22	2.16	1.37	1.69	2.98	1.85	2.15
Planters (Con)	1.92	2.02	2.3	4.16	1.52	4.1	5.13

Individual Estimates

	Model/Rank:		I = 1	I = 2		I = 3			I = 4			
	Weight on individual:		100.0%	81.2%	18.8%	62.8%	31.4%	5.8%	73.5%	24.1%	2.4%	0.02%
	Product	Logit Sj	i = 1	i = 1	i = 2	i = 1	i = 2	i = 3	i = 1	i = 2	i = 3	i = 4
SALTY SNACKS	Snyders (Con)	2.21	0.70	0.64	0.70	0.00	0.00	2.17	0.00	2.38	0.00	0.23
	Cheetos	2.52	0.50	0.00	0.00	0.24	0.25	0.80	0.31	0.00	0.00	0.00
	Ruffles (Con)	0.98	1.88	0.98	4.82	0.00	2.53	4.84	0.03	5.05	2.54	0.00
	Dorito Nacho	2.05	0.98	0.90	1.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rold Gold (Con)	0.94	1.86	2.35	0.00	2.50	0.13	1.40	0.65	1.95	0.14	4.41
	Baked (Con)	1.99	2.08	2.49	0.40	1.36	0.00	3.94	2.35	3.76	0.00	0.43
	Salty Other (Con)	2.78	0.22	0.29	0.00	0.05	0.00	0.59	0.00	0.71	0.00	0.30
	Sun Chip	1.81	4.76	0.00	22.96	0.00	19.29	5.19	0.09	5.71	18.70	0.00
	Cheez-It	1.77	1.47	1.06	2.67	0.00	0.91	3.90	0.00	4.00	0.93	0.43
	Jays (Con)	1.48	0.17	0.23	0.00	0.04	0.00	0.47	0.00	0.57	0.00	0.24
	Frito	2.03	1.41	1.28	1.65	0.00	0.00	4.55	0.00	4.70	0.00	0.04
	FritoLay (Con)	1.49	1.71	1.57	1.94	0.15	0.36	4.49	0.53	4.50	0.45	0.34
	Smartfood	1.51	0.56	0.67	0.00	0.31	0.16	1.02	0.00	1.17	0.18	0.69
	Lays	1.45	0.54	0.56	0.31	0.00	0.00	1.62	0.00	1.68	0.00	0.30
	Cheetos Flamin	0.96	0.55	0.50	0.55	0.00	0.00	1.83	0.00	1.92	0.03	0.00
	Dorito Blazin	1.45	1.47	0.01	6.47	0.00	5.77	1.82	0.00	2.06	5.62	0.00
	Popcorn (Con)	2.06	0.51	0.63	0.00	0.61	0.33	0.34	0.22	0.54	0.36	0.80
	Ritz Bits	0.51	0.16	0.22	0.00	0.17	0.00	0.23	0.00	0.32	0.03	0.26
CHOCOLATE CANDY	M&M Peanut	3.21	4.78	6.46	0.00	8.95	0.00	1.18	16.65	0.00	0.00	0.00
	Snickers	3.53	6.08	8.00	0.00	9.75	0.70	0.00	14.91	0.00	0.11	0.13
	Twix Caramel	2.29	5.19	7.32	0.00	11.03	0.00	0.00	4.04	0.00	0.00	18.60
	Raisinets	1.47	1.47	2.03	0.00	2.67	0.00	0.21	2.74	0.03	0.05	2.06
	M&M Milk Choc	1.80	3.63	4.90	0.00	6.47	0.00	0.67	5.31	0.36	0.04	7.24
	Choc Mars (Con)	2.13	1.03	1.46	0.00	2.03	0.00	0.11	0.00	0.31	0.00	4.69
	Reeses PB Cups	1.68	1.62	2.98	0.00	4.68	0.00	0.36	2.95	0.25	0.00	7.38
	Butterfinger	1.10	2.72	3.21	0.80	3.69	1.13	1.67	2.06	1.73	1.16	5.62
	Choc Herhsey (Con)	1.22	2.87	1.58	7.20	1.64	5.92	2.92	0.91	3.23	5.77	2.53

	Model/Rank:		I = 1	I = 2		I = 3			I = 4			
	Weight on individual:		100.0%	81.2%	18.8%	62.8%	31.4%	5.8%	73.5%	24.1%	2.4%	0.02%
	Product	Logit Sj	i = 1	i = 1	i = 2	i = 1	i = 2	i = 3	i = 1	i = 2	i = 3	i = 4
NONCHOC. CANDY	Skittles Original	1.03	0.12	0.18	0.00	0.03	0.00	0.36	0.00	0.43	0.00	0.17
	Nonchoc Other (Con)	1.06	0.41	0.59	0.00	0.65	0.00	0.32	0.00	0.41	0.00	1.54
	Twizzlers	1.66	1.16	1.20	0.86	0.90	0.59	1.71	1.96	1.41	0.66	0.00
COOKIES	ZAnimal Cracker	1.90	0.29	0.35	0.14	0.33	0.39	0.00	0.16	0.00	0.00	0.00
	CC Fam Amos	1.58	1.57	0.00	3.66	0.04	26.42	0.00	0.23	0.00	28.73	0.00
	Ruger Wafer (Con)	1.60	0.54	0.68	0.00	0.46	0.00	0.94	0.00	1.07	0.00	1.31
	Grandmas CC	1.15	0.84	0.46	2.07	0.34	2.21	0.89	0.47	0.92	2.25	0.05
	Rasbry Knotts	0.68	1.10	0.47	3.18	0.45	2.89	1.12	0.76	1.19	2.87	0.00
	Choc Fam Amos	0.91	1.35	1.09	2.12	1.47	2.65	0.38	1.19	0.52	2.63	1.35
	Nabisco (Con)	1.23	1.44	1.22	2.04	1.24	2.28	1.11	0.99	1.20	2.21	1.44
PASTRY	Pop-Tarts (Con)	2.42	0.27	0.39	0.00	0.34	0.00	0.36	0.00	0.46	0.00	0.87
	Rice K Treats	0.85	2.25	2.64	0.80	2.06	0.16	3.22	2.78	3.03	0.24	1.59
OTHER	Nature Valley (Con)	2.13	1.42	1.47	1.02	0.42	0.00	3.54	1.63	3.22	0.00	0.00
	Planters (Con)	1.63	4.81	3.51	9.13	4.39	8.99	2.79	4.91	2.79	8.74	3.10
	KarNuts (Con)	1.65	1.25	1.68	0.00	1.71	0.00	1.05	2.51	0.87	0.01	0.36
	Farleys Fruit Snax	0.99	0.58	0.57	0.45	0.04	0.00	1.69	0.16	1.69	0.00	0.08
	Cherry Fruit Snax	0.52	0.09	0.14	0.00	0.05	0.00	0.25	0.00	0.30	0.00	0.12
	Cliff (Con)	3.91	1.03	1.29	0.00	1.30	0.51	0.67	0.62	0.82	0.48	2.03
	Outside Good	25.34	28.58	29.75	22.86	27.43	15.42	33.28	27.87	32.77	15.06	29.27

Extensions and Conclusion

Extensions

- What about (exogenous) price or quality changes?
  Expression for $D_{j\to k}$ changes slightly if just quality index $\beta_i = \beta$ like ξ_j
- Want to add covariates?
  Straightforward to run an IV regression:

$$\log \widehat{s}_{ij} - \log \widehat{s}_{i0} = V_{ij} - V_{i0} = f_i(x_j) - \alpha_i p_j + \xi_j$$

  Test how much we lose using only a basis in $f_i(x_j)$.
- A real world vending experiment with 8 product removals – here we don't see the entire $\mathcal{D}_{j\to k}$ and must complete it.
- Optimal Experimentation: Which product is most informative about $\mathcal{D}$?
  - Not quite a theory: probably high centrality ones (!)

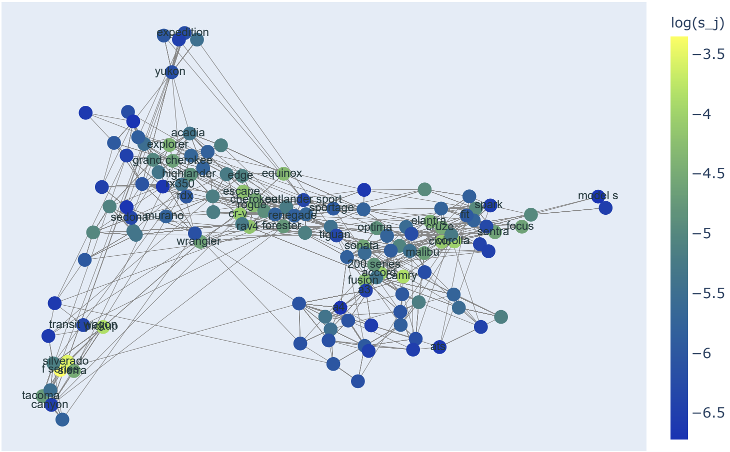
Diversion as a Network

The matrix $\mathcal{D}$ has some useful properties:

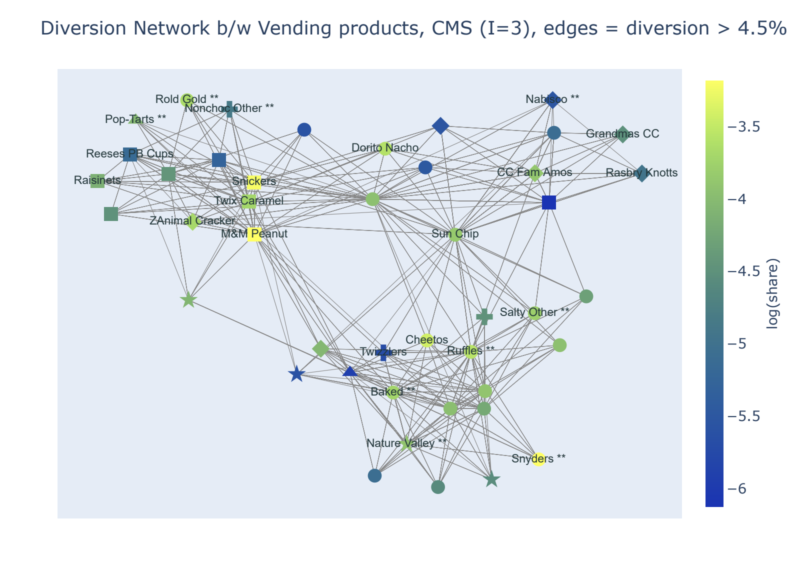
- We know that $\mathcal{D}_{jk} \in [0, 1]$.
- Each row $\sum_k \mathcal{D}_{j \to k} = 1$.
- $\mathcal{D}$ looks like a transition matrix with a network structure
- We can represented $\mathcal{D}_{jk}$ as a weighted directed graph.
- I've already used some graph algorithms to sort the rows/columns of $\mathcal{D}$.

Network Structure of Cars: Raw Data

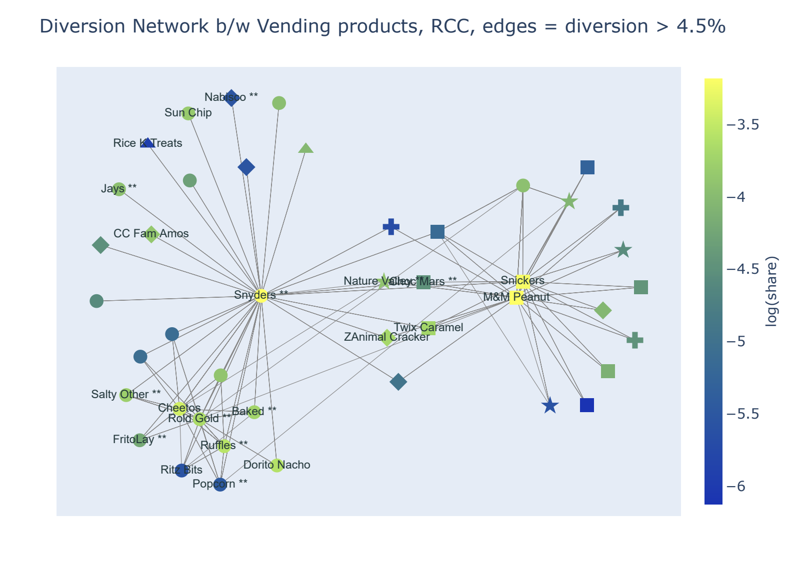
Diversion Network b/w 2015 Cars, edges = diversion > 4.5%



Network Structure of Snacks: Our Estimates



Network Structure of Snacks: RCC



Diversion as a Network: Work in Progress

What else can we do with this representation?

- Can we apply graph clustering to minimize number of “cuts” to segment into distinct markets/categories?
- Some products (nodes) have high centrality but low share? Are these products that are likely to provide “discipline” in merger settings?
  - Relates to measures of centrality / eigenvalues.
  - Cross elasticities are not a well-behaved network.
- Which centrality measure (early results seem to suggest Bonacich).

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 (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"?
- Which products are important for completing substitution patterns?