Estimating Preferences and Substitution Patterns from Second-Choice Data Alone

Chris Conlon (NYU Stern & NBER), Julie Holland Mortimer (University of Virginia & NBER), and Paul Sarkis (Analysis Group)

January, 2025

Motivation

The too-many-parameters problem: a complete substitution matrix has J^2 parameters, but aggregate data have only J observations. In order to predict substitution, we need to restrict consumers' choice problem (reduce dimensionality). Two options:

1. In product space, impose structure on the form of the underlying utility functions

$$\ln q_{jt} = \gamma_j + \sum_k \alpha_{jk} \ln p_{kt} + e_{jt}$$

- ▶ Examples: log log or AIDS (Deaton and Muellbauer (1980 AER))
- Often with multi-level structure so that $\alpha_{jk} = 0$ for certain pairs.

2. In characteristic space, assume substitution patterns are functions of product characteristics

$$u_{ij} = \beta_i x_j + \xi_j + \varepsilon_{ij}.$$

Mixed Logit: Explain substitution patterns using observed characteristics

- Often 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 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 substitution patterns.

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

Even the most flexible and data-rich random-coefficients models suffer from three main deficiencies:

- 1. Never quite enough substitution to best substitutes
- 2. Everything looks a bit too much like plain logit (substitution proportional to share)
- 3. Substitution patterns are only as good as your characteristics → if you want extreme substitution patterns you need extreme characteristics.

For competition policy enforcement:

- Agencies may have (possibly limited) data on substitution, but not much else
 - Second-choice surveys (e.g., Sainsbury's/Asda, Microsoft/Activision)
 - Customer switching or win/loss data (e.g., cell phone companies).
 - Natural or field experiments (hospital closures; product removals)
- These may not be the objects we want to use as substitution patterns.
 - Often we want to understand *ceteris paribus* responses to small changes in price.

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

Can we construct a low-rank approximation to substitution patterns in product space?

- ▶ Idea: avoid the too-many-problem by directly restricting the rank of substitution matrix
- ▶ Trick: re-cast substitution in terms of second-choice probabilities
- ▶ We show this leads to a convenient semi-parametric representation.

- Product characteristics do not accurately capture substitution across products.
- ▶ We lack sufficient variation in prices, other covariates, to estimate demand system.
- ▶ We need to estimate substitution patterns across all products but have data on only a subset:
 - ▶ shares of cell phone providers, but number porting/win-loss data for a subset (merging parties)
 - ▶ second-choice surveys (UK CMA: Sainsbury's/Asda, Microsoft/Activision, Amazon/Deliveroo)
 - ► aggregate market shares and a subset of second-choice data (microBLP (2004); Grieco, Murry, Yurukoglou (2024)).

Suppose we observe some aggregate shares $\mathcal{S} = [\mathcal{S}_1, \dots, \mathcal{S}_J]$, and number porting data \mathcal{D}^T

| | VZ | ATT | ТМо | S | Other | | |
|-------------------|------------------|-----|------|------|-------|-------|---------------------------------|
| | 0 | ? | 0.30 | 0.30 | ? \ | VZ | $\left\lceil 0.35 \right\rceil$ |
| | ? | 0 | 0.45 | 0.15 | 0 | ATT | 0.30 |
| $\mathcal{D}^T =$ | ? | ? | 0 | 0.45 | ? | TMo , | 0.20 = S |
| | ? | ? | 0.20 | 0 | ? | S | 0.10 |
| | $\backslash ? =$ | ? | 0.05 | 0.10 | 0 / | Other | $\lfloor 0.05 \rfloor$ |

Can we fill in the missing elements and simulate the merger?

Example of Second-Choice Survey Data, 318 Cars and Light Trucks

Second-Choice Data



Image of Camille Jordan (1838-1922)



Review of Diversion Ratios

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

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

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

 $\bullet D_{j \to k} \equiv \frac{\partial q_k}{\partial p_j} / \left| \frac{\partial q_j}{\partial p_j} \right|.$

▶ Can also write as $D_{j \to k} \equiv \frac{\epsilon_{kj}}{|\epsilon_{jj}|} \cdot \frac{q_j}{q_k}$

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:

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

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 $z_j \rightarrow z'_j$ (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 $\mathcal J$

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

And let $dF_i = \pi_i$ be weight on each type *i* (Monte Carlo/Quadrature/etc):

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

Different interventions give different weights (but don't change individual $D_{j \rightarrow k,i}$).

| | $w_{ij}(x) \propto$ |
|--|--|
| second-choice data | $s_{ij}(x)$ |
| price change $\frac{\partial}{\partial p_j}$ | $s_{ij}(x)(1-s_{ij}(x))\cdot \alpha_i $ |
| characteristic change $\frac{\partial}{\partial x_i}$ | $s_{ij}(x)(1-s_{ij}(x))\cdot \beta_i $ |
| small quality change $\frac{\partial}{\partial \xi_i}$ | $s_{ij}(x)(1-s_{ij}(x))$ |
| finite price change $w_i(p_j,p_j^\prime,x)$ | $ s_{ij}(p'_j, x) - s_{ij}(p_j, x) $ |
| finite quality change $w_i(\xi_j,\xi_j',x)$ | $ s_{ij}(\xi'_j, x) - s_{ij}(\xi_j, x) $ |
| willingness to pay (WTP) | $rac{s_{ij}(x)}{ lpha_i \cdot s_{i0}(x)}$ |

Data and the Problem of Rank Reduction

- **1**. From a parametric model: estimate demand and compute $\left[\frac{\partial q_j}{\partial p_j}(\mathbf{p})\right]^{-1} \frac{\partial q_k}{\partial p_j}(\mathbf{p})$
- 2. Farrell Shapiro (2011) hoped for info gathered "in an email" or "normal course of business"
- 3. From observed "Win-Loss" data (slightly different weights)
- 4. Randomized choice sets
 - Chris and I did this in vending machines.
 - Many examples of online search results/product listings
- 5. From surveys of second choices (e.g. stated preference)

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

$$S_j = \mathbb{P}(\text{chooses } j \in \mathcal{J})$$

 $\mathcal{D}_{j \to k} = \mathbb{P}(\text{chooses } k \in \mathcal{J} \setminus \{j\} \mid \text{chooses } j \in \mathcal{J})$

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 of Second-Choice $\mathcal{D}_{i \rightarrow k}$, 318 Cars and Light Trucks, MaritzCX

Second-Choice Data



Consider a low-rank approximation to substitution patterns using data on second choices.

- 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.
- If $A \sim B$ and $B \sim C$ then either $A \sim C$ or we need to increase the rank!

Setting up the Estimator

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) ; integrate out over distribution of V_{ij} :

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

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

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:

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

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

- Individual *i*'s share for each choice given by $\mathbf{s_i} = [s_{i0}, s_{i1}, \dots, s_{iJ}]$.
- 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:

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

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

$$\mathbf{D} = \operatorname{diag}(\mathbf{s})^{-1} \cdot \left(\sum_{i=1}^{I} \pi_{i} \cdot \begin{bmatrix} & | \\ & \mathbf{s}_{\mathbf{i}} \\ & | \end{bmatrix} \cdot \begin{bmatrix} & - & \frac{\mathbf{s}_{\mathbf{i}}}{1 - \mathbf{s}_{\mathbf{i}}} & - & \end{bmatrix} \right)$$

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

 $\operatorname{Fix} \operatorname{rank}(\mathbf{D}(\mathbf{S}, \pi)) = I$, and for each choice of I solve:

$$\min_{(\mathbf{S},\pi) \ge 0} \left\| \mathcal{P}_{\Omega}(\mathcal{D} - \mathbf{D}(\mathbf{S},\pi)) \right\|_{\ell_{2}} + \lambda \left\| \mathcal{S} - \mathbf{S} \,\pi \right\|_{\ell_{2}} \text{ with } \left\| \pi \right\|_{\ell_{1}} \le 1, \quad \left\| \mathbf{s}_{\mathbf{i}} \right\|_{\ell_{1}} \le 1.$$

• Goal: estimate s_i (choice probabilities) and corresponding weights π_i (Finite Mixture)

- Not convex, but not very difficult either.
- Constraints: Choice probabilities s_{ij} sum to one, type weights π_i sum to one.
 - ℓ_1 constraints lead to sparsity.
- ▶ Idea: Control the rank by limiting *I* directly
 - Use cross validation to select # of types I and Lagrange multiplier λ .
- Matrix completion: We can construct estimates of $D(S, \pi)$ including elements of $\mathcal{P}_{\overline{\Omega}}$.

- Absent any constraints from discrete choice (and first-choice probabilities) we know the solution is similar to the Camille Jordan problem
 - ▶ Take the first *I* singular values from the SVD (as in Camille Jordan example)
- ▶ The Nuclear Norm of a matrix $\|D\|_*$ is the sum of its singular values and provides a "continuous approximation" to its rank.
 - ▶ Measure how complicated our second-choice data/parametric models are.

Comparisons

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

$$\begin{split} \min_{\pi_i \ge 0} \sum_j \left(\mathcal{S}_j - \sum_i \pi_i \cdot \hat{s}_{ij}(\hat{\beta}_i) \right)^2 \quad \text{subject to} \quad \sum_i \pi_i = 1 \\ \hat{s}_{ij}(\hat{\beta}_i) = \frac{e^{\hat{\beta}_i x_j}}{1 + \sum_{j'} e^{\hat{\beta}_i x'_j}} \end{split}$$

- 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 950/1000 simulated consumers).
- > Data requirement: characteristics that vary across markets.
- Fix grid of $\hat{\beta}_i$ (and thus \hat{s}_{ij}); search over π_i .

See Heiss, Hetzenecker, Osterhaus (JoE 2022).

- 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.

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

- ▶ Use second choices from hospital closures (natural disasters) to compare models.
- ▶ Requires individual data grouped into bins, plus characteristics.
- Fix π_i 's, search for β_g (and thus $s_{g(i)}$).

Most similar to what we're doing conceptually.

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

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

► We work in product space (no need for characteristics), using the summing-up constraints to enforce that the model is consistent with discrete choice.

Monte Carlo

- Fit (i) nested logit, (ii) RC logit to data on vending machines (Conlon and Mortimer JPE, 2021).
- Generate fake sales and second-choices 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 second-choice probabilities.
 - ► MAD: Median $(|\mathcal{D}_{j\to k} \hat{D}_{j\to k}|)$ for $(j,k) \in P_{\overline{\Omega}}$ (Validation).

• RMSE:
$$\sqrt{\frac{1}{n}\sum_{(j,k)\in P_{\overline{\Omega}}\}} |\mathcal{D}_{j\to k} - \hat{D}_{j\to k}|^2}$$

Monte Carlo: DGP is Nested Logit

Cross-validation Results - Nested Logit DGP



29/57

Monte Carlo: DGP is RC on chars (Interpolation!)

Cross-validation Results - RC Logit DGP



30/57

Application to Autos Data

- Subset of data from Grieco, Murry and Yurukoglu (QJE 2023).
- 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.
- How: Split sample into P_{Ω} and $P_{\overline{\Omega}}$

MaritzCX Survey data (318 Cars and Light Trucks)

Second-Choice Data



Cross Validation: Model Selection



In-Sample Performance



34/57

Analysis of Consumer Weights



Consumer Weights Concentration

Comparison of Implied Diversion

Data

GMY



CMS (I=15)





Profiles of Types (Rank 15)



Top Substitutes: Ford F-Series

| Model | Raw | Logit | CMS I=15 | CMS I=30 | GMY | |
|---------------------|-------|-------|----------|----------|-------|---|
| Ram Pickup | 24.59 | 0.88 | 21.46 | 22.23 | 19.40 | |
| Gmc Sierra | 20.29 | 0.61 | 14.97 | 21.92 | 17.27 | |
| Chevrolet Silverado | 15.62 | 0.78 | 13.41 | 19.63 | 33.62 | |
| Toyota Tundra | 12.98 | 0.55 | 16.32 | 12.79 | 2.29 | |
| Toyota Tacoma | 6.31 | 0.76 | 3.39 | 3.13 | 2.83 | |
| Chevrolet Colorado | 4.64 | 0.63 | 3.22 | 2.86 | 2.87 | |
| Gmc Canyon | 2.30 | 0.30 | 0.76 | 1.38 | 1.02 | |
| Nissan Frontier | 1.63 | 0.43 | 0.92 | 1.69 | 0.61 | |
| Jeep Wrangler | 1.59 | 0.69 | 1.33 | 0.94 | 0.06 | |
| Nissan Titan | 0.70 | 0.05 | 1.18 | 1.17 | 0.18 | |
| Ford Explorer | 0.63 | 0.38 | 0.16 | 0.14 | 0.71 | |
| | | | | | | 1 |

Top Substitutes: Honda Odyssey

| Model | Raw | Logit | CMS I=15 | CMS I=30 | GMY |
|-------------------------|-------|-------|----------|----------|-------|
| Toyota Sienna | 44.51 | 0.50 | 41.02 | 41.43 | 28.34 |
| Chrysler Town & Country | 11.27 | 0.44 | 13.70 | 13.66 | 6.90 |
| Dodge Caravan | 8.67 | 0.61 | 12.11 | 12.90 | 7.04 |
| Kia Sedona | 8.38 | 0.18 | 7.22 | 7.22 | 7.59 |
| Mazda Mazda5 | 2.02 | 0.19 | 0.07 | 0.00 | 0.01 |
| Nissan Quest | 2.02 | 0.16 | 3.15 | 2.56 | 2.43 |
| Honda Pilot | 1.73 | 0.29 | 1.40 | 1.64 | 0.59 |
| Chevrolet Traverse | 1.73 | 1.33 | 1.50 | 1.29 | 0.18 |
| Toyota Highlander | 1.45 | 1.17 | 1.30 | 1.24 | 0.53 |
| Gmc Acadia | 1.16 | 0.63 | 1.36 | 1.36 | 0.15 |
| Ford Flex | 1.16 | 0.10 | 0.86 | 1.02 | 0.03 |

Top Substitutes: Mercedes-Benz Sprinter Van

| Model | Raw | Logit | CMS I=15 | CMS I=30 | GMY |
|-----------------------------|-------|-------|----------|----------|------|
| Ford Transit Wagon | 66.67 | 0.58 | 18.18 | 51.72 | 0.04 |
| Ram Promaster | 16.67 | 0.08 | 0.00 | 0.00 | 1.76 |
| Ford Transit Connect | 8.33 | 0.66 | 22.12 | 0.04 | 0.01 |
| Nissan Nv | 8.33 | 0.47 | 15.75 | 30.18 | 6.52 |
| Mini Cooper | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 |
| Volkswagen Beetle li Cabrio | 0.00 | 0.28 | 0.02 | 0.00 | 0.00 |
| Audi A5 | 0.00 | 0.31 | 0.04 | 0.00 | 0.02 |
| Mazda Mx-5 Miata | 0.00 | 0.33 | 0.04 | 0.01 | 0.00 |
| Audi S5 | 0.00 | 0.18 | 0.04 | 0.00 | 0.00 |
| Porsche Boxster | 0.00 | 0.18 | 0.80 | 0.01 | 0.00 |
| Volkswagen Eos | 0.00 | 0.06 | 0.00 | 0.00 | 0.00 |

Second Stage

- So far: Estimation of $\widehat{\mathbf{S}}$ and \widehat{s}_{ij} , as well as $\widehat{\pi}$.
- Challenge: Cannot directly calculate price elasticities and consumer welfare.
- ▶ Solution: Follow a two-stage approach inspired by existing literature.
- Second Stage: Recovery of price sensitivity coefficients.
 - Assumption: Utility $u_{ij} = V_{ij} + \varepsilon_{ij}$ with ε_{ij} as an IID Type I extreme value error term.
 - Key Equation:

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

Then either calibrate β^p_i = ∂V_{ij}/∂p_j using (1) observed own-price elasticities or
 (2) observed price-cost margins.

Own-Price Elasticity Calibration

- Approach 1: Calibrate $\beta_i^p = \frac{\partial V_{ij}}{\partial p_i}$ using observed own-price elasticities.
 - Sources of observed elasticities: Quasi-experiment, other studies, or simpler demand systems on observational data.
- Minimum Distance Estimator:

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

- Requirements:
 - Observe at least as many elasticities as types I.
 - Simplifies if $\beta_i^p = \beta^p$, where a single or average elasticity identifies β^p .
- Empirical Example: Estimating β_i^p using own-elasticities estimated from GMY.

Implied Price Coefficients



43/57

Model Fit



44/57

Price-Cost Margin Calibration

- Approach 2: Calibrate $\beta_i^p = \frac{\partial V_{ij}}{\partial p_i}$ using observed price-cost margins.
 - Ideally at product or firm level.
- Construct Δ matrix of demand derivatives, ${\cal H}$ ownership matrix.
- P-C margin is then:

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

Minimum Distance Estimator:

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

• Empirical Example: Estimating β_i^p using price-cost margins estimates from GMY.

Implied Price Coefficients



Estimated Price coefficients (CMS I=15),

Model Fit



47/57

Vending Example

- ▶ 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)

| Product | \mathcal{S}_{j} | $\mathcal{D}_{j \to k}$ | 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 | 7.10 |
| 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 | 0.50 | 5.22 | 1.22 | 1.71 | 2.49 | 3.75 | 4.20 |
| Raisinets | 1.60 | 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 | 0.00 | 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 | 2.30 | 4.16 | 1.52 | 4.10 | 5.13 |

Individual Estimates

| | Mod | del/Rank: | l = 1 | 1= | 2 | | I = 3 | | 1 = 4 | | | |
|-------|--------------------|------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | Weight on ir | ndividual: | 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 |
| | 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 |
| S | Sun Chip | 1.81 | 4.76 | 0.00 | 22.96 | 0.00 | 19.29 | 5.19 | 0.09 | 5.71 | 18.70 | 0.00 |
| MAC | Cheez-It | 1.77 | 1.47 | 1.06 | 2.67 | 0.00 | 0.91 | 3.90 | 0.00 | 4.00 | 0.93 | 0.43 |
| LTV S | Jays (Con) | 1.48 | 0.17 | 0.23 | 0.00 | 0.04 | 0.00 | 0.47 | 0.00 | 0.57 | 0.00 | 0.24 |
| 3 | 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 |
| | 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 |
| CAN | Raisinets | 1.47 | 1.47 | 2.03 | 0.00 | 2.67 | 0.00 | 0.21 | 2.74 | 0.03 | 0.05 | 2.06 |
| LATE | 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 |
| 8 | 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 | 1= | 2 | | I = 3 | | | 1 = 4 | | | | |
|--------------------------------------|---------------------|------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--|--|
| | Weight on ir | ndividual: | 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 | | |
| OTHER PASTRY COOKIES NOWCHOC COOKIES | Skittles Original | 1.03 | 0.12 | 0.18 | 0.00 | 0.03 | 0.00 | 0.36 | 0.00 | 0.43 | 0.00 | 0.17 | | |
| AND' | 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 | | |
| NOI | Twizzlers | 1.66 | 1.16 | 1.20 | 0.86 | 0.90 | 0.59 | 1.71 | 1.96 | 1.41 | 0.66 | 0.00 | | |
| | 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 | | |
| 8 | 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 | | |
| OKI | Grandmas CC | 1.15 | 0.84 | 0.46 | 2.07 | 0.34 | 2.21 | 0.89 | 0.47 | 0.92 | 2.25 | 0.05 | | |
| 8 | 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 | | |
| TRY | 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 | | |
| BAS | 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 | | |
| | 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 | | |
| 8 | KarNuts (Con) | 1.65 | 1.25 | 1.68 | 0.00 | 1.71 | 0.00 | 1.05 | 2.51 | 0.87 | 0.01 | 0.36 | | |
| OTP | 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

- ► What about (exogenous) price or quality changes? Expression for D_{j→k} changes slightly
- Want to add covariates?
 Straightforward to run an IV regression:

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

Test how much we lose using only a basis in $f_i(x_j)$.

► Can we estimate q_0 instead of assuming it: $S_j = \frac{q_j}{q_0 + \sum_{k \in \mathcal{J} \setminus \{0\}} q_k}$? Yes.

- Removing multiple choices at once
 - Conceptually easy: $\frac{s_{ik}}{1-s_{ij}-s_{i\ell}}$
- Observing substitution at more consolidated level (ie: Make vs. Model)
- 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 *D*?
 - Not quite a theory: probably high centrality ones

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

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

Network Structure of Snacks: Our Estimates

Diversion Network b/w Vending products, CMS (I=3), edges = diversion > 4.5%



Network Structure of Snacks: RCC

Diversion Network b/w Vending products, RCC, edges = diversion > 4.5%



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).

Wrapping Up

- > Things work well and move away from logit's insufficient sparsity
- Inference: What is asymptotic experiment?
 - Minimum Distance?
 - Taking $J \rightarrow \infty$? (like Berry Linton Pakes 2004)
 - \blacktriangleright Taking number of surveyed columns/individuals in ${\cal D}$ to be complete?
- Identification
 - We know that I = 1 (logit) is probably identified
 - We know that $I \gg rank(\mathcal{D})$ is probably not
 - For I < rank(D) what else do we need to guarantee uniqueness? (Matrix factorization problems are not always unique)
- How much information on second choices is "enough"?
- Explore which products matter for completing substitution patterns.