Empirical Properties of Diversion Ratios

Christopher T. Conlon and Julie Holland Mortimer

July 18, 2018

NYU Stern and Boston College & NBER

What are Diversion Ratios?

Horizontal Merger Guidelines (2010 rev.)

In some cases, the Agencies may seek to quantify the extent of direct competition between a product sold by one merging firm and a second product sold by the other merging firm by estimating the diversion ratio from the first product to the second product. The diversion ratio is the fraction of unit sales lost by the first product due to an increase in its price that would be diverted to the second product. Diversion ratios between products sold by one merging firm and products sold by the other merging firm can be very informative for assessing unilateral price effects, with higher diversion ratios indicating a greater likelihood of such effects. Diversion ratios between products sold by merging firms and those sold by non-merging firms have at most secondary predictive value.

This time with equations

Raise price of good j. People leave. What fraction of leavers switch to k?

$$D_{jk} = \frac{\frac{\partial q_k}{\partial p_j}}{\left|\frac{\partial q_j}{\partial p_j}\right|}$$

It's one of the best ways economists have to characterize competition among sellers.

- High Diversion: Close Substitutes $\rightarrow$ Mergers more likely to increase prices.
- Very low diversion $\rightarrow$ products may not be in the same market. (ie: Katz & Shapiro)
- Demand Derivatives NOT elasticities.
- No equilibrium responses.

Where do Diversion Ratios come from?

In Theory

Consider Bertrand FOC's for single-product firm j, buys k:

$$
\begin{aligned}
\arg\max_{p_j} \quad & (p_j - c_j) \cdot q_j(p_j, p_{-j}) + (p_k - c_k) \cdot q_k(p_j, p_{-j}) \\
0 &= q_j + (p_j - c_j) \cdot \frac{\partial q_j}{\partial p_j} + (p_k - c_k) \cdot \frac{\partial q_k}{\partial p_j} \\
p_j &= -q_j / \frac{\partial q_j}{\partial p_j} + c_j + (p_k - c_k) \cdot \underbrace{\frac{\partial q_k}{\partial p_j} / - \frac{\partial q_j}{\partial p_j}}_{D_{jk}} \\
p_j &= \underbrace{\frac{\epsilon_{jj}}{\epsilon_{jj} + 1}}_{\text{Lerner Markup}} \left[c_j - \overbrace{\underbrace{c_j \cdot e_j}_{\text{efficiency}} + \underbrace{(p_k - c_k) \cdot D_{jk}(p_j, p_{-j})}_{\text{opp cost}}}^{UPP} \right]
\end{aligned}
$$

Caveat: UPP, Partial Merger, Full Merger.

In Practice

1. Calculated from an estimated demand system (ratio of estimated cross-price to own-price demand derivatives)
2. Consumer surveys (what would you buy if not this?)
3. Obtained in ‘course of business’ (sales reps, internal reviews)

Antitrust authorities may prefer different measures in different settings. Are they concerned about:

- Small but widespread price hikes?
- Product discontinuations or changes to availability?

Is it sufficient to rely on data from merging firms only?

- Do we need diversion to other products in the ‘market’ or other functions of market-level data?
- Discrete-choice demand models imply that ‘aggregate diversion’ (including to an outside good) sums to one.

Literature

Goal: connect two literatures with little overlap.

Applied theory motivates use of diversion ratios:

- Farrell & Shapiro (2010)
- Is diversion informative about price effects? (Don't address.)
  - Shapiro (1995), Werden (1996), Werden and Froeb (2006), Carlton (2010), Schmalensee (2009), Willig (2011), Jaffe and Weyl(2013).
  - Miller, Remer, Ryan and Sheu (2012), Cheung (2011) - price effects, prediction errors depend on nature of competition for non-merging firms.
  - Reynolds & Walters (2008) - consumer surveys

Applied econometrics articulates estimation challenges:

- Angrist, Graddy and Imbens (2000) shows how a cost shock identifies a LATE for price elasticity with one product.
- Berry, Leveinsohn and Pakes (2004) examines economic content of second-choice data.
- Use of experimental techniques vis-a-vis structural methods.
  - See 2010 JEP collection on Leamer (1983): Angrist and Pischke, Nevo and Whinston, Heckman, Leamer, Keane, Sims, Stock, Einav and Levin.

Gameplan

Approach: Learn the empirical properties of diversion ratios.

- Some econometric properties using program evaluation framework
  - Diversion as a matrix function
  - LATE vs. ATE. vs. MTE
  - Bias/Variance Tradeoff
  - Second Choice Data
- Standard data (Nevo (2000)): discrete-choice demand
  - Estimate the MTE and ATE and compare
  - Check against the plain logit
- Large-scale experiment: exogenous product removals
  - Estimate ATE only; no need for parametric demand models
  - Observing data from all products in the market (not just those in the merger) is important
  - The "summing up" constraint may be more important for identification than a parametric distribution on error terms.
  - Diversion ratios may help to identify good candidate products for divestiture.

A Simple Insight...

Diversion has treatment effects interpretation

Treatment	"not purchasing j"
Outcome	fraction of j consumers who switch to product k
Treated group	consumers who would have purchased j at pre-merger price, but do not purchase at a higher price

Heterogeneity: Individuals who leave j after a \$0.01 price increase differ in their taste for k from those who leave after \$1, \$100, \$10,000 price increases.

Start with the Wald Estimator

Consider an experiment designed to measure diversion, where everything else is held fixed and p_j is exogenously increased by Δp_j:

$$D_{jk}(p_j, p_{-j}) = \left| \frac{q_k(p_j + \Delta p_j, p_{-j}) - q_k(p_j, p_{-j})}{q_j(p_j + \Delta p_j, p_{-j}) - q_j(p_j, p_{-j})} \right| = \frac{\int_{p_j^0}^{p_j^0 + \Delta p_j} \frac{\partial q_k(p_j, p_{-j})}{\partial p_j} \partial p_j}{\Delta q_j}$$

Re-write as Local Average Treatment Effect

$$\widehat{D_{jk}}^{LATE} = \frac{1}{\Delta q_j} \int_{p_j^0}^{p_j^0+\Delta p_j} \underbrace{\frac{\partial q_k}{\partial q_j}}_{D_{jk}(p_j, p_{-j})} \left| \frac{\partial q_j(p_j, p_{-j})}{\partial p_j} \right| \partial p_j$$

- $\widehat{D_{jk}}^{LATE}$ is a Local Average Treatment Effect (LATE).
  - Identified from finite price changes (simulated or actual).
  - For any finite price increase, we measure a weighted average of the diversion function, where the weights are the lost sales of j: $w(\mathbf{p}) = \frac{1}{\Delta q_j} \frac{\partial q_j(p_j, p_{-j})}{\partial p_j}$

Re-write as Local Average Treatment Effect

$$\widehat{D_{jk}}^{LATE} = \frac{1}{\Delta q_j} \int_{p_j^0}^{p_j^0+\Delta p_j} \underbrace{\frac{\partial q_k}{\partial q_j}}_{D_{jk}(p_j, p_{-j})} \left| \frac{\partial q_j(p_j, p_{-j})}{\partial p_j} \right| \partial p_j$$

- $\widehat{D_{jk}}^{LATE}$ is a Local Average Treatment Effect (LATE).
    - Identified from finite price changes (simulated or actual).
    - For any finite price increase, we measure a weighted average of the diversion function, where the weights are the lost sales of j: $w(\mathbf{p}) = \frac{1}{\Delta q_j} \frac{\partial q_j(p_j, p_{-j})}{\partial p_j}$
- Let $\widehat{D_{jk}}^{ATE}$ denote Average Treatment Effect (ATE) when everyone is treated.
    - Δp_j increases to choke price: $Q_j(p_j^0 + \Delta p_j, p_{-j}) = 0$.
    - Interpretation as second-choice data.

The Nonparametric Object: MTE

Re-writing:

$$\widehat{D_{jk}}^{LATE} = \frac{1}{\Delta q_j} \int_{p_j^0}^{p_j^0 + \Delta p_j} D_{jk}(p_j, p_{-j}) \left| \frac{\partial q_j(p_j, p_{-j})}{\partial p_j} \right| \partial p_j$$

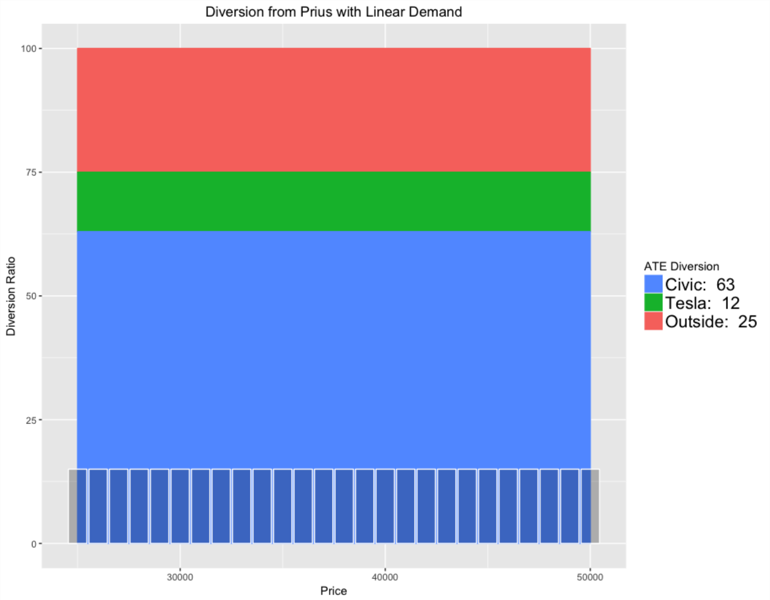
- Diversion, $D_{jk}(p_j, p_{-j})$, is a Marginal Treatment Effect (MTE) in the language of Heckman and Vytlacil (1999).
- It is a function. Actually a matrix valued function.
- It is not identified non-parametrically from a single price increase.

Various Treatment Effects

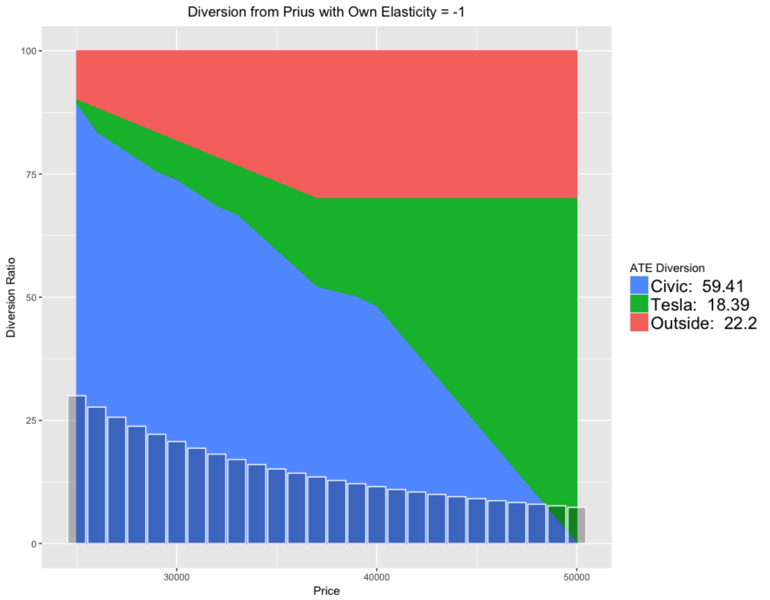
- Determine what different measures of diversion identify.
  - Finite price increase $\rightarrow$ local average treatment effect (LATE)
  - Product removal (treating everyone) $\rightarrow$ average treatment effect (ATE)
  - A nonparametric function of p_j $\rightarrow$ marginal treatment effect (MTE)
  - Constant diversion: three measures coincide (Theory/Empirics)

But... How do the weights work? An illustration.

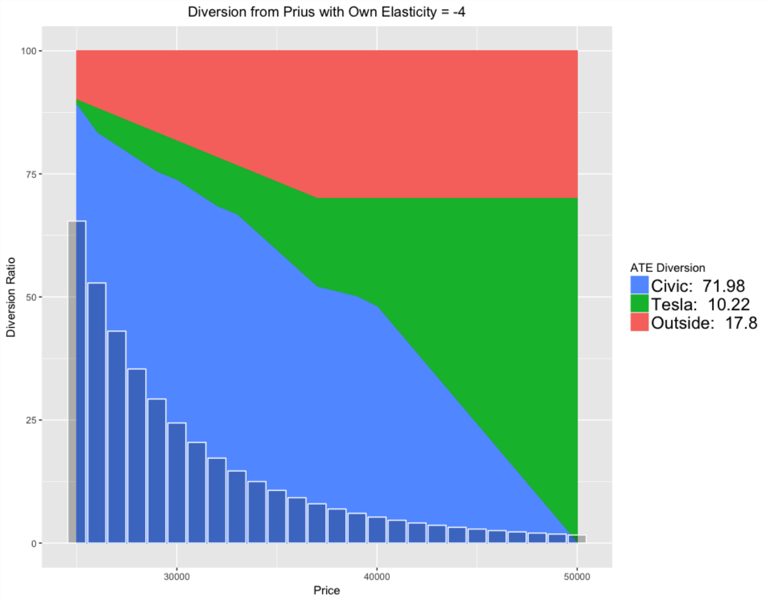
Thought Experiment – Linear Demand for a Toyota Prius



Thought Experiment – Inelastic Demand for a Prius



Thought Experiment – Elastic Demand for a Prius



Bias of Estimator

How far apart are $D_{jk}(\mathbf{p^0})$ and $\widehat{D_{jk}}$ when we increase price by Δp_j?

$$
\begin{aligned}
q_k(\mathbf{p} + \Delta p_j) &\approx q_k(\mathbf{p}) + \frac{\partial q_k}{\partial p_j}\Delta p_j + \frac{\partial^2 q_k}{\partial p_j^2}(\Delta p_j)^2 + O((\Delta p_j)^3) \\
\frac{q_k(\mathbf{p} + \Delta p_j) - q_k(\mathbf{p})}{\Delta p_j} &\approx \frac{\partial q_k}{\partial p_j} + \frac{\partial^2 q_k}{\partial p_j^2}\Delta p_j + O(\Delta p_j)^2 \\
Bias(\widehat{D_{jk}} - D_{jk}(\mathbf{p^0})) &\approx -\frac{D_{jk}\frac{\partial^2 q_j}{\partial p_j^2} + \frac{\partial^2 q_k}{\partial p_j^2}}{\frac{\partial q_j}{\partial p_j} + \frac{\partial^2 q_j}{\partial p_j^2}\Delta p_j}\Delta p_j
\end{aligned}
$$

- The downside of a large change Δp_j is that the approximation of demand at $\mathbf{p^0}$ is less accurate and depends on the curvature of the demand function.
- There are two demand models for which $Bias \equiv 0$ (constant treatment effects):

Bias of Estimator

How far apart are $D_{jk}(\mathbf{p^0})$ and $\widehat{D_{jk}}$ when we increase price by Δp_j?

$$\begin{aligned}
q_k(\mathbf{p}+\Delta p_j) &\approx q_k(\mathbf{p}) + \frac{\partial q_k}{\partial p_j}\Delta p_j + \frac{\partial^2 q_k}{\partial p_j^2}(\Delta p_j)^2 + O((\Delta p_j)^3) \\
\frac{q_k(\mathbf{p}+\Delta p_j) - q_k(\mathbf{p})}{\Delta p_j} &\approx \frac{\partial q_k}{\partial p_j} + \frac{\partial^2 q_k}{\partial p_j^2}\Delta p_j + O(\Delta p_j)^2 \\
Bias(\widehat{D_{jk}} - D_{jk}(\mathbf{p^0})) &\approx -\frac{D_{jk}\frac{\partial^2 q_j}{\partial p_j^2} + \frac{\partial^2 q_k}{\partial p_j^2}}{\frac{\partial q_j}{\partial p_j} + \frac{\partial^2 q_j}{\partial p_j^2}\Delta p_j}\Delta p_j
\end{aligned}$$

- The downside of a large change Δp_j is that the approximation of demand at $\mathbf{p^0}$ is less accurate and depends on the curvature of the demand function.
- There are two demand models for which $Bias \equiv 0$ (constant treatment effects): linear demand and plain IIA logit.

Variance of Estimator

$$Var(\widehat{D_{jk}}) \quad \approx \quad Var\left(\frac{\Delta q_k}{|\Delta q_j|}\right) \approx \frac{D_{jk}(1-D_{jk})}{\Delta q_j} \approx \frac{D_{jk}(1-D_{jk})}{\left|\frac{\partial q_j}{\partial p_j}\right| \Delta p_j}$$

Variance is a problem when:

- $\left|\frac{\partial q_j}{\partial p_j}\right|$ is small (inelastic demand $\rightarrow$ market power/ when we may be most worried about mergers).
- $\Delta p_j \approx 0$ (when price change is small).
- Exacerbated by variation in (q_j, q_k) unrelated to the exogenous price change (stochastic demand).

Bias-Variance tradeoff

- Precise measure of $\widehat{D_{jk}}^{ATE}$ or $\widehat{D_{jk}}^{LATE}$ for a large Δp_j vs.
- Noisy measure of $D_{jk}(\mathbf{p^0})$

Which Diversion Ratio do we want?

Small price changes, and variance is small: LATE for small Δp

- Airline prices, consumer goods and services in large markets.
- Examples of other merger cases here
- Anhauser-Busch/InBev (beer prices)

Large price change/discontinued products, large variance: ATE

- Airline routes, hospital networks
- Merger of two European machinery firms
- Data storage (Dell and EMC merger.)
- Snack foods (Planter's Cheez Balls after Kraft merger.)

Nevo (2000): MTE vs. ATE

Nevo (2000) Application

- Data from Nevo (2000): $T = 94$ markets, $J = 24$ brands.
- RTE cereal (e.g., Kellogg's and General Mills merger)

$$u_{ijt} = d_j + x_{jt} \underbrace{\left(\overline{\beta} + \Sigma \cdot \nu_i + \Pi \cdot d_{it}\right)}_{\beta_{it}} + \Delta\xi_{jt} + \varepsilon_{ijt}$$

- Features a large amount of preference heterogeneity, especially with respect to the price sensitivity β_{it}^{price}
- Estimated coefficient on price is distributed:

$$\beta_{it}^{price} \sim N\left(\text{-63} + 588 \cdot \text{income}_{it} \text{ - } 30 \cdot \text{inc}_{it}^{2} + 11 \cdot \text{I[child]}_{it}, \sigma{=}3.3\right)$$

Nevo (2000), cont.

Define:

$$MTE = \frac{\frac{\partial s_k}{\partial p_j}}{\left|\frac{\partial s_j}{\partial p_j}\right|}, \quad ATE = \frac{s_k(A \setminus j) - s_k(A)}{|s_j(A \setminus j) - s_j(A)|}, \quad Logit = \frac{s_k(A)}{1 - s_j(A)}$$

- Compare $MTE(\mathbf{p_0})$ to ATE
- Compare $MTE(\mathbf{p_0})$ to Logit (Constant diversion, $\propto$ to share.)

Nevo (2000), cont.

Three Measures of Diversion

	MTE	*ATE*	*Logit*
	Best Substitute		
Med(D_{jk})	13.26	13.54	9.05
Mean(D_{jk})	15.11	15.62	10.04
% Agree with MTE		89.98	58.38
	Outside Good		
Med(D_{j0})	35.30	32.40	54.43
Mean(D_{j0})	36.90	33.78	53.46

The first panel reports diversion to each product-market pair's best substitute. The second panel reports diversion to the outside good.

Nevo (2000), cont.

% Difference in Diversion Measures: y vs. $x = \log(\widehat{D^{MTE}(\mathbf{p_0})})$

	med$(y-x)$	mean$(y-x)$	med$(\lvert y-x\rvert)$	mean$(\lvert y-x\rvert)$	std$(\lvert y-x\rvert)$
			Best Substitutes		
ATE	2.56	3.24	6.00	7.61	7.04
Logit	-44.19	-42.88	44.92	47.77	28.63
			All Products		
ATE	5.78	8.30	8.29	12.13	12.02
Logit	-35.90	-25.92	49.48	53.27	34.56
			Outside Good		
ATE	-7.93	-8.89	7.94	9.08	6.77
Logit	39.22	39.20	39.22	40.60	22.05

Table compares ATE and Logit measures of diversion to the MTE measure.

The first panel reports differences for each product-market pair's best substitute.

The second panel averages across all possible substitutes.

The third panel provides comparisons to the MTE diversion for the outside good.

Lessons from Nevo (2000)

- MTE vs. ATE measures are not hugely different.
- ATE tends to predict slightly more inside substitution and less outside substitution. Why?
- They both rely on sum of diversion $= 1$.
- Imposing proportional substitution (Logit) looks terrible.

Experimental ATE

Vending Application with Exogenous Product Removal

- Conlon and Mortimer (2010, 2017): exogenous product removals.
- Multiple treatments; 14,000 - 19,000 exposed consumers in each. ($\approx$ 900 treated)
- Probably not worried about bias. Why?
- Main challenge in this exercise is going to be variance.
- Still IO economists, a little economic theory goes a long way.

Vending Application with Exogenous Product Removal

- Conlon and Mortimer (2010, 2017): exogenous product removals.
- Multiple treatments; 14,000 - 19,000 exposed consumers in each. ($\approx$ 900 treated)
- Probably not worried about bias. Why?
- Main challenge in this exercise is going to be variance.
- Still IO economists, a little economic theory goes a long way.

Application uses vending machines. Lots of mergers!

- Kellogg/Keebler, then Kellogg/Pringles
- Kraft/Phillip Morris, then Kraft/Nabisco, then Kraft/Heinz
- Kraft became Mondelez
- Hershey/Reese's, then Hershey/Cadbury
- Mondelez/Hershey scuttled by Hershey Trust

Manufacturer Share and HHI's by Category and Total

Manufacturer:	Category:			
	Salty Snack	Cookie	Confection	Total
PepsiCo	78.82	9.00	0.00	37.81
Mars	0.00	0.00	58.79	25.07
Hershey	0.00	0.00	30.40	12.96
Nestle	0.00	0.00	10.81	4.61
Kellogg's	7.75	76.94	0.00	11.78
Nabisco	0.00	14.06	0.00	1.49
General Mills	5.29	0.00	0.00	2.47
Snyder's	1.47	0.00	0.00	0.69
ConAgra	1.42	0.00	0.00	0.67
TGIFriday	5.25	0.00	0.00	2.46
Total	100.00	100.00	100.00	100.00
HHI	6332.02	6198.67	4497.54	2401.41

IRM Brandshare FY 2006 and Frito-Lay Direct Sales For Vending Machines, Heartland Region, 50 top products.
(http://www.vending.com/Vending_Affiliates/Pepsico/Heartland_Sales_Data)

Experimental Design 1

Setting:

- Mark Vend, a mid-size vending operator
- 60 snack machines in office buildings in downtown Chicago
- Spread across 5 client locations/6 sites
- Aggregate sales to client-week level
- "White collar" customer base, fairly stable demand over time
- Very few "natural stock-outs" occur.

Experiment:

- We focus on four experiments from two manufacturers
  - Mars: Snickers, M&M Peanut
  - Kellogg: Animal Crackers, Famous Amos Cookies

Experimental Design 2

Removals are equivalent to increasing p_j to the choke price, for which $q_j = 0$. No one has access to product j.

- For each run, we removed the focal product(s) for about 2.5-3 weeks from all machines at each site.
- Machines are visited 3 times on avg. during each removal.
- Data were collected from January, 2006 - February, 2009.
- Experimental periods range from June 2007 to September 2008.
- Interventions were run during the months of May - October.
- Poster-card announcements were included at the front of all empty product columns.
    - "This product is temporarily unavailable.
      We apologize for any inconvenience."

Summary Statistics for Experimental Treatments

	Control Period	Snickers	Zoo Animal Crackers	Famous Amos	M&M Peanut
# Machines	66	62	62	62	56
# Weeks	160	6	5	4	6
# Machine-Weeks	8,525	190	161	167	223
# Products	76	67	65	67	66
Total Sales	700,404.0	16,232.5	14,394.0	13,910.5	19,005.2
—Per Week	4,377.5	2,705.4	2,878.8	3,477.6	3,167.5
—Per Mach-Week	82.2	85.4	89.4	83.3	85.2
Total Focal Sales		42,047.8	26,113.2	21,578.4	44,026.3
—Per Week		262.8	163.2	134.0	273.5
—Per Mach-Week		4.9	3.1	2.5	5.2

Variance Problem Part 1

Variance Problem Part 2

Manufacturer	Product	Control Mean	Treatment Mean	Treatment Quantile
			Snickers Removal	
Mars	M&M Peanut	309.8	472.5	100.0
Pepsi	Rold Gold (Con)	158.9	331.9	91.2
Mars	Twix Caramel	169.0	294.1	100.0
Pepsi	Cheeto LSS	248.6	260.7	61.6
Snyders	Snyders (Con)	210.2	241.6	52.8
Kellogg	Animal Cracker	183.1	233.7	96.8
Kraft	Planters (Con)	161.1	218.8	96.0
	Total	4892.1	5357.9	74.4
			Zoo Animal Crackers Removal	
Mars	M&M Peanut	309.7	420.3	99.2
Mars	Snickers	301.3	385.1	94.4
Pepsi	Rold Gold (Con)	158.9	342.4	92.0
Snyders	Snyders (Con)	210.3	263.0	67.2
Pepsi	Cheeto LSS	248.6	263.0	66.4
Mars	Twix Caramel	169.1	235.0	99.2
Pepsi	Baked Chips (Con)	169.6	219.7	89.6
	Total	4892.2	5608.6	89.6

Estimating $\widehat{D_{jk}}^{ATE}$

Goal: Construct a non-parametric estimate for ATE guided by economic theory.

Assumption 1: Valid Control Weeks For a machine-week observation to be included as a control for $q_{k,t}$ it must: (a) have product k available; (b) be from the same vending machine; (c) not be included in any of our treatments.

Assumption 2: Substitutes Removing a product does not increase overall sales, or decrease overall sales by more than q_j.

Assumption 3: Unit Interval $D_{jk} \in [0, 1]$.

Assumption 4: Unit Simplex $D_{j\cdot} \in \Delta$. Sum of diversion to all substitutes (including outside good) is one.

Using just Assumptions (1)+(2)

Problem is that distribution of demand shocks ξ are quite different

$$f(\xi | T = 1) \neq f(\xi | T = 0)$$

Old IO problem, life would be better if we saw the ξ.

We develop a form of a matching estimator.

- Choose control weeks, s so that $Q_s - Q_t \in [0, q_{js}]$.
- Have to be careful because q_{js} leads to selection bias (Weeks with unusually high q_{js} are more likely to be included as controls).
- Solve this by using fitted value from regression of q_{js} on Q_s.

$$S_t = \{s : Q_s - Q_t \in [0, \widehat{b_0} + \widehat{b_1} Q_s]\}$$

Matching is not enough

Assumptions (1) and (2) solve a few problems:

- (Not shown) Restricting to cases where k is available, other conditions of Assn (1) help address unobserved heterogeneity across time and machines.
- Substitution to the outside good improves with addition of Assn (2).

But problems remain:

- Nearly half of products exhibit negative diversion ratios.
- Many products do not have 'reasonable' diversion.
- The 'best substitute' is driven by a small sample size.
- Diversion to the top 5 products exceeds 200%.

Need some baseball (Efron and Morris 78)

Table 1.1: Batting averages $z_i = \hat{\mu}_i^{(\mathrm{MLE})}$ for 18 major league players early in the 1970 season; μ_i values are averages over the remainder of the season. The James–Stein estimates $\hat{\mu}_i^{(\mathrm{JS})}$ (1.35) based on the z_i values provide much more accurate overall predictions for the μ_i values. (By coincidence, $\hat{\mu}_i$ and μ_i both average 0.265; the average of $\hat{\mu}_i^{(\mathrm{JS})}$ must equal that of $\hat{\mu}_i^{(\mathrm{MLE})}$.)

Name	hits/AB	$\hat{\mu}_i^{(\mathrm{MLE})}$	μ_i	$\hat{\mu}_i^{(\mathrm{JS})}$
Clemente	18/45	.400	**.346**	.294
F Robinson	17/45	.378	**.298**	.289
F Howard	16/45	.356	**.276**	.285
Johnstone	15/45	.333	**.222**	.280
Berry	14/45	.311	**.273**	.275
Spencer	14/45	.311	**.270**	.275
Kessinger	13/45	.289	**.263**	.270
L Alvarado	12/45	.267	**.210**	.266
Santo	11/45	.244	**.269**	.261
Swoboda	11/45	.244	**.230**	.261
Unser	10/45	.222	**.264**	.256
Williams	10/45	.222	**.256**	.256
Scott	10/45	.222	**.303**	.256
Petrocelli	10/45	.222	**.264**	.256
E Rodriguez	10/45	.222	**.226**	.256
Campaneris	9/45	.200	**.286**	.252
Munson	8/45	.178	**.316**	.247
Alvis	7/45	.156	**.200**	.242
Grand Average		.265	**.265**	.265

Assumption (3): Using a Beta-Binomial Prior

How to restrict $D_{jk} \in [0, 1]$?

$$\Delta q_k | \Delta q_j, D_{jk} \quad \sim \quad Bin(n = \Delta q_j, p = D_{jk})$$

Assumption (3): Using a Beta-Binomial Prior

How to restrict $D_{jk} \in [0,1]$?

$$\begin{aligned}
\Delta q_k | \Delta q_j, D_{jk} &\sim Bin(n = \Delta q_j, p = D_{jk}) \\
D_{jk} | \beta_1, \beta_2 &\sim Beta(\beta_1, \beta_2) \\
E[D_{jk} | \beta_1, \beta_2, \Delta q_j, \Delta q_k] &= \frac{\beta_1 + \Delta q_k}{\beta_1 + \beta_2 + \Delta q_j} \\
\mu_{jk} &= \frac{\beta_1}{\underbrace{\beta_1 + \beta_2}_{m_{jk}}}, \quad \lambda = \frac{m_{jk}}{m_{jk} + \Delta q_j} \\
\widehat{D_{jk}} &= \lambda \cdot \mu_{jk} + (1 - \lambda) \frac{\widehat{\Delta q_k}}{\widehat{\Delta q_j}}
\end{aligned}$$

μ_{jk} is prior mean; m_{jk} is no. pseudo-obs; λ weights our prior mean.
When we have a lot of experimental obs, prior receives little weight.

Assumption (4): Using a Dirichlet Prior

How do restrict $D_{j\cdot} \in \Delta$?

- Same idea as before, but use Dirichlet Prior.
- Acts like pseudo-observations from the multinomial distribution.
- If we had same number of treated observations for each substitute we would have conjugacy/closed form (We don't).
- We use $m = 3.05$ pseudo-observations for the logit prior (extremely weak).
    - Add $m = 1.1$ pseudo-observations from uniform (keeps things away from zero).
    - Location of prior is largely irrelevant.
- Estimator is still technically non-parametric. Why?

Shrinkage Estimator, Intuition

- We "shrink" towards the prior mean when we have experimental estimates that are imprecise.
- Idea is very simple: when we have lots of data, use the experimental measure.
- When data are scarce: put more weight on the prior/model-based measure.
  - In practice: FTC/DOJ tend to assume diversion proportional to marketshare
  - Use plain logit (could also use more complicated model)
  - Logit sets the mean of the prior as: $\mu_{jk} = \frac{s_k}{1-s_j}$

Results: Mars

Firm	Product	# Weeks	Δq_k	Δq_j	$\frac{\Delta q_k}{\Delta q_j}$	Beta(J)	Beta(300)	Dirichlet(4.15)
		Snickers Removal						
Mars	M&M Peanut	176	375.52	-954.30	39.35	37.04	30.80	18.40
Mars	Twix Caramel	134	289.60	-702.39	41.23	37.86	29.49	15.88
Pepsi	Rold Gold (Con)	174	161.37	-900.11	17.93	16.84	13.95	7.54
Nestle	Butterfinger	61	72.95	-362.82	20.11	17.07	11.19	4.45
Mars	M&M Milk Chocolate	97	71.76	-457.36	15.69	13.83	9.85	4.14
Kraft	Planters (Con)	136	78.01	-759.87	10.27	9.57	7.80	3.81
Kellogg	Zoo Animal Cracker	177	65.72	-970.22	6.77	6.48	5.68	2.92
Pepsi	Sun Chip	159	45.30	-866.09	5.23	4.98	4.33	2.07
Hershey	Choc Hershey (Con)	41	29.78	-179.57	16.58	12.17	6.30	2.01
	Outside Good	180	460.89	-970.22	47.50			23.12
		M&M Peanut Removal						
Mars	Snickers	218	296.58	-1239.29	23.93	22.90	19.91	16.47
Mars	Twix Caramel	176	110.93	-1014.32	10.94	10.39	8.88	6.76
Mars	M&M Milk Chocolate	99	73.47	-529.58	13.87	12.46	9.18	6.26
Nestle	Raisinets	181	71.82	-1001.14	7.17	6.82	5.82	4.37
Kraft	Planters (Con)	190	61.42	-1046.10	5.87	5.62	4.90	3.60
Hershey	Twizzlers	62	32.98	-332.99	9.90	8.32	5.32	3.35
Kellogg	Rice Krispies Treats	46	22.37	-220.17	10.16	7.90	4.43	2.51
Pepsi	Frito	160	37.25	-902.42	4.13	3.95	3.47	2.37
	Outside Good	218	606.18	-1238.49	48.95			36.35

Results: Kellogg's

Firm	Product	# Weeks	Δq_k	Δq_j	$\frac{\Delta q_k}{\Delta q_j}$	Beta(J)	Beta(300)	Dirichlet(4.15)
		Animal Crackers Removal						
Pepsi	Rold Gold (Con)	132	114.39	-440.80	25.95	22.90	16.21	9.89
Mars	Snickers	145	92.44	-483.63	19.11	17.26	13.04	7.58
Mars	M&M Peanut	142	77.72	-469.44	16.55	14.98	11.43	6.47
Kellogg	CC Famous Amos	144	66.18	-478.20	13.84	12.40	9.15	5.39
Pepsi	Baked Chips (Con)	134	62.55	-447.60	13.97	12.46	9.13	5.27
Mars	Twix Caramel	110	50.17	-338.97	14.80	12.75	8.74	4.58
Sherwood	Ruger Wafer (Con)	119	48.20	-368.65	13.07	11.28	7.63	4.28
Hershey	Choc Herhsey (Con)	30	33.60	-132.57	25.34	17.14	7.86	3.81
Kellogg	Rice Krispies Treats	13	23.52	-37.80	62.22	23.24	7.16	2.99
Kar's Nuts	Kar Sweet&Salty Mix	95	30.06	-334.50	8.99	7.72	5.27	2.73
Misc	Popcorn (Con)	56	25.72	-226.89	11.34	8.92	5.08	2.61
Kraft	Planters (Con)	114	28.05	-380.25	7.38	6.53	4.78	2.43
Mars	M&M Plain	73	22.67	-295.07	7.68	6.47	4.26	2.15
	Outside Good	145	240.52	-482.91	49.81			21.98
		Famous Amos Removal						
Pepsi	Sun Chip	139	143.60	-355.68	40.37	34.39	22.66	15.75
Kraft	Planters (Con)	121	82.11	-332.61	24.69	20.89	13.68	8.75
Hershey	Choc Hershey (Con)	38	48.60	-66.84	72.72	36.93	13.36	7.18
Pepsi	Frito	119	49.88	-313.21	15.93	13.44	8.85	5.32
Misc	Rasbry Knotts	133	46.62	-345.38	13.50	11.45	7.49	4.81
Pepsi	Grandmas Choc Chip	95	39.99	-259.21	15.43	12.51	7.62	4.49
Pepsi	Dorito Buffalo Ranch	72	38.11	-224.24	17.00	13.28	7.53	4.43
Pepsi	Chs PB Frito Cracker	34	26.87	-83.65	32.13	18.16	7.14	3.74
Kellogg	Choc Sandwich FA	57	27.97	-122.04	22.91	15.06	6.84	3.69
Pepsi	Rold Gold (Con)	147	32.62	-392.22	8.32	7.40	5.54	3.19
Kraft	Oreo Thin Crisps	29	20.73	-43.29	47.89	19.20	6.12	3.05
Mars	Combos (Con)	98	23.56	-274.54	8.58	7.03	4.34	2.61
	Outside Good	156	192.90	-399.12	48.33			20.95

Comparison of Assumptions: Snickers Experiment

	Total	Assn 1	Assn 2	Assn 3 ($m = K$)	Assn 4 (m = 4.15)
Products with $D_{jk} < 0$	51	24	26	0	0
Products with $0 \leq D_{jk} \leq 10$	51	13	15	43	48
Products with $10 \leq D_{jk} \leq 20$	51	5	5	5	2
Products with $D_{jk} > 20$	51	9	5	3	1
Sum of all positive D_{jk}s	51	402.84	301.95	265.41	98.72
Sum of all negative D_{jk}s	51	-238.90	-239.07	0.00	0.00

Note: Table includes only products for which there were at least 50 sales of the focal product in control weeks, on average.

Posterior Distribution for Diversion Assn (1)-(4) $m = 4.15$

Manuf	Product	Mean	2.5^{th} Quantile	25^{th} Quantile	50^{th} Quantile	75^{th} Quantile	97.5^{th} Quantile
Snickers Removal							
Nestle	Nonchoc Nestle (Con)	0.67	0.31	0.51	0.65	0.81	1.17
Mars	Twix Caramel	15.88	14.28	15.32	15.88	16.45	17.53
Mars	M&M Peanut	18.40	16.79	17.83	18.39	18.95	20.02
Kellogg	Rice Krispies Treats	1.30	0.78	1.09	1.28	1.49	1.95
Nestle	Butterfinger	4.45	3.53	4.10	4.43	4.78	5.48
Mars	Choc Mars (Con)	0.44	0.16	0.31	0.42	0.55	0.85
Pepsi	Rold Gold (Con)	7.54	6.49	7.15	7.53	7.92	8.69
	Outside Good	23.12	21.34	22.50	23.11	23.73	24.91
Zoo Animal Crackers Removal							
Kellogg	Rice Krispies Treats	2.99	1.93	2.56	2.95	3.36	4.28
Misc	Salty United (Con)	1.25	0.61	0.97	1.21	1.49	2.12
Kraft	100 Cal Oreo Thin Crisps	1.85	1.04	1.51	1.81	2.14	2.88
Pepsi	Rold Gold (Con)	9.89	8.24	9.30	9.88	10.46	11.66
Hershey	Choc Herhsey (Con)	3.81	2.66	3.35	3.77	4.22	5.17
Misc	Hostess Pastry	1.80	1.02	1.47	1.76	2.08	2.79
Kraft	100 Cal Chse Nips Crisps	1.10	0.51	0.83	1.06	1.32	1.91
	Outside Good	21.98	19.64	21.15	21.96	22.78	24.43

Divestiture: Kellogg's-Mars Merger

	Treated Machine Weeks	No Prior	Assn 3 Diversion ($m = K$)	Assn 4 Diversion ($m = 4.15$)
Snickers to Kellogg's Products				
Zoo Animal Cracker	177	6.77	6.53 (5.13; 8.09)	2.92 (2.27; 3.65)
CC Famous Amos	180	4.61	4.46 (3.29; 5.79)	1.99 (1.40; 2.69)
Choc Sandwich FA	69	8.39	7.31 (5.18; 9.77)	1.98 (1.45; 2.59)
Rice Krispies Treats	17	26.68	14.18 (8.66; 20.66)	1.31 (0.79; 1.96)
Cheez-It Original SS	150	0.26	0.35 (0.07; 0.82)	0.10 (0.01; 0.27)
Pop-Tarts*	162	-4.28	0.10 (0.00; 0.38)	0.00 (0.00; 0.02)
Total (to Kellogg's)		42.44	32.93 (26.54; 40.04)	8.30 (7.14; 9.56)
Outside Good	180	47.50	46.19 (43.17; 49.27)	23.19 (21.39; 24.99)
Peanut M&M to Kellogg's Products				
Rice Krispies Treats	46	10.16	7.87 (5.12; 11.12)	2.74 (1.73; 3.95)
CC Famous Amos	215	0.27	0.30 (0.10; 0.66)	0.17 (0.04; 0.40)
Cheez-It Original SS	188	-4.81	0.09 (0.00; 0.31)	0.00 (0.00; 0.03)
Zoo Animal Cracker	218	-2.62	0.10 (0.01; 0.32)	0.00 (0.00; 0.04)
Pop-Tarts*	191	-1.80	0.07 (0.00; 0.27)	0.00 (0.00; 0.03)
Choc Sandwich FA	70	-0.89	0.05 (0.00; 0.34)	0.00 (0.00; 0.03)
Total (to Kellogg's)		0.30	8.46 (5.71; 11.74)	2.92 (1.89; 4.15)
Outside Good	218	48.95	47.85 (45.26; 50.33)	37.62 (35.45; 39.82)

- For manufacturers: $p = .45, c = .15$ is a good approximation $\rightarrow$ would require marginal cost reductions $2\times$ diversion.
- Adding up constraint substantially reduces overall diversion.
- Without adding up constraint looks like Rice Krispies Treats would be likely divestiture.

Divestiture: Kellogg's-Mars Merger

	Treated Machine Weeks	No Prior	Assn 3 Diversion ($m = K$)	Assn 4 Diversion ($m = 4.15$)
Zoo Animal Crackers to Mars' Products				
Snickers	145	19.11	17.28	7.59
			(14.42; 20.48)	(6.17; 9.14)
M&M Peanut	142	16.55	15.24	6.47
			(12.23; 18.53)	(5.12; 7.93)
Twix Caramel	110	14.80	12.90	4.58
			(9.82; 16.58)	(3.44; 5.87)
M&M Milk Chocolate	73	7.68	6.46	2.16
			(4.30; 9.26)	(1.38; 3.11)
Milky Way	9	22.62	7.81	0.86
			(3.32; 14.51)	(0.35; 1.61)
Combos (Con)	95	-8.79	0.07	0.00
			(0.00; 0.52)	(0.00; 0.03)
Non-chocolate candy*	114	-20.12	0.17	0.00
			(0.00; 0.74)	(0.00; 0.05)
Total (to Mars)		51.86	60.04	21.68
			(52.26; 68.89)	(19.30; 24.12)
Outside Good	145	49.81	46.99	22.02
			(43.09; 51.01)	(19.71; 24.42)
Chocolate Chip Famous Amos to Mars' Products				
Combos (Con)	98	8.58	7.23	2.61
			(4.74; 10.22)	(1.71; 3.71)
Milky Way	26	19.47	10.63	1.94
			(5.92; 16.49)	(1.07; 3.08)
Twix Caramel	121	-9.85	0.31	0.01
			(0.01; 1.07)	(0.00; 0.07)
Snickers	156	-14.62	0.43	0.01
			(0.05; 1.20)	(0.00; 0.09)
M&M Peanut	153	-16.26	0.45	0.01
			(0.06; 1.26)	(0.00; 0.09)
Non-chocolate candy*	124	-17.46	0.20	0.00
			(0.00; 0.82)	(0.00; 0.05)
M&M Milk Chocolate	89	-19.77	0.21	0.00
			(0.00; 1.01)	(0.00; 0.05)
3-Musketeers	82	-63.25	0.17	0.00
			(0.00; 0.89)	(0.00; 0.04)
Total (to Mars)		-113.16	19.67	4.59
			(14.07; 26.30)	(3.28; 6.13)
Outside Good	156	48.33	45.26	20.97
			(40.69; 49.80)	(18.43; 23.61)

- Without adding up constraint MilkyWay looks like divestiture candidate.
- Adding up constraint reduces diversion by factor of 3.
- Much more concern about Animal Cracker price increase than Famous Amos.

Final Plea: A little advocacy...

Isn't this:

	Snickers	M&M Peanut	Animal Cracker	Famous Amos
Mars	39.4	31.3	21.7	4.6
Kellogg	8.3	2.9	10.2	4.1
Pepsi	12.1	4.2	18.8	39.7
Kraft	4.0	4.4	5.8	11.8
Hershey	2.9	7.8	3.8	8.9
Nestle	5.7	6.7	2.0	2.0
Misc	4.4	5.1	15.6	7.9
Outside Good	23.2	37.6	22.0	21.0

Better than cross elasticities $[.002, .003, .001, \ldots]$.

Conclusion

- What are we measuring? What do we want?
    - Widespread but small price increases (MTE)
    - Product discontinuations/ 2nd choices (ATE)
    - How curvy is demand?
- My Checklist
    - How does outside good diversion look?
    - Enough flexibility across substitutes/columns?
    - Enough flexibility across focal products/rows?
- For Practicioners
    - Plain Logit seems really bad (no parameters!)
    - Observing all substitutes in category seems helpful.