NBER WORKING PAPER SERIES

AN EXPERIMENTAL APPROACH TO MERGER EVALUATION

Christopher T. Conlon
Julie Holland Mortimer

Working Paper 19703
http://www.nber.org/papers/w19703

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2013

We thank Dennis Carlton, Joe Farrell, Dan O'Brien, and participants at the FTC Microeconomics Conference and 'I.O. Fest' at UC Berkeley for comments. Tom Gole, Adam Kapor, Joonwoo Park, and Sharon Traiberman provided exceptional research assistance. Financial support for this research was generously provided through NSF grant SES-0617896. Any remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

An Experimental Approach to Merger Evaluation
Christopher T. Conlon and Julie Holland Mortimer
NBER Working Paper No. 19703
December 2013
JEL No. L0,L4,L44

ABSTRACT

The 2010 Department of Justice and Federal Trade Commission Horizontal Merger Guidelines lay out a new standard for assessing proposed mergers in markets with differentiated products. This new standard is based on a measure of ``upward pricing pressure," (UPP) and the calculation of a ``gross upward pricing pressure index" (GUPPI) in turn relies on a ``diversion ratio," which measures the fraction of consumers of one product that switch to another product when the price of the first product increases. One way to calculate a diversion ratio is to estimate own- and cross-price elasticities. An alternative (and more direct) way to gain insight into diversion is to exogenously remove a product from the market and observe the set of products to which consumers actually switch. In the past, economists have rarely had the ability to experiment in this way, but more recently, the growth of digital and online markets, combined with enhanced IT, has improved our ability to conduct such experiments. In this paper, we analyze the snack food market, in which mergers and acquisitions have been especially active in recent years. We exogenously remove six top-selling products (either singly or in pairs) from vending machines and analyze subsequent changes in consumers' purchasing patterns, firm profits, diversion ratios, and upward pricing pressure. Using both nonparametric analyses and structural demand estimation, we find significant diversion to remaining products. Both diversion and the implied upward pricing pressure differ significantly across manufacturers, and we identify cases in which the GUPPI would imply increased regulatory scrutiny of a proposed merger.

Christopher T. Conlon
Department of Economics
Columbia University
420 W. 118th St.
New York City, NY 10027
cc3264@columbia.edu

Julie Holland Mortimer
Department of Economics
Boston College
140 Commonwealth Avenue
Chestnut Hill, MA 02467
and NBER
julie.mortimer.2@bc.edu

1 Introduction

Since 1982, one of the primary tools in evaluating the potential anticompetitive effects of horizontal mergers has been the Herfindahl-Hirschman Index (HHI) which measures both the level of concentration in a particular market, as well as how a proposed merger would change the level of concentration. The HHI relates market shares to markups when firms are engaged in a symmetric Cournot game, but because HHI requires market shares as inputs, a practical challenge has been defining the relevant market in terms of both geography and competitors. The 2007 merger between *Whole Foods* and *Wild Oats* highlighted the problems with the black-and-white distinction of market definitions. In that case, the FTC argued that the proposed merger was essentially a merger to monopoly in the market for "premium organic groceries," while Whole Foods argued they faced competition from traditional grocery stores, and thus the merger represented little change in concentration.

In 2010, the Department of Justice (DOJ) and the Federal Trade Commission (FTC) released a major update to the Horizontal Merger Guidelines, which shifts the focus away from traditional concentration measures like HHI, and towards methods that better account for product differentiation and the closeness of competition, utilizing intuition from the differentiated products Bertrand framework. Rather than use a full structural equilibrium simulation of post-merger prices and quantities such as in Nevo (2001), as an initial screen, the guidelines instead hold all post-merger quantities fixed, as well as the prices of all goods outside the merger, and consider the unilateral effects of the merger on the prices of the merged entity's products. This exercise produces the two key measures: Upward Pricing Pressure (UPP) and Generalized Upward Pricing Pressure Index (GUPPI). Both of these depend on the prices and costs of merged products, and are monotonic functions of a "Diversion Ratio" between the products of the two merged firms. In general, one expects prices and profit margins to be directly observed, leaving only the diversion ratio to be estimated.[1] In this sense, the diversion ratio serves as a sufficient statistic to determine whether or not a proposed merger is likely to increase prices (and be contested by the antitrust authorities).

This sentiment is captured in the Guidelines' definition of the diversion ratio:

> *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*

[1] Prices and margins may be observed by antitrust authorities, even if academic researchers do not observe them. Under the Hart-Scott-Rodino Antitrust Improvements Act of 1976, price and margin data are provided in pre-merger filings.

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.

There is a growing literature that examines the potential advantages and disadvantages of the use of diversion ratios as the primary input into merger evaluation. Many of these advantages are discussed in Farrell and Shapiro (2010), and similar to those found in the literature on sufficient statistics (Chetty 2009). One benefit is that the diversion ratio does not require data from all firms in an industry, but rather only those firms engaged in a potential merger. Another benefit is that the diversion ratio need not be measured with reference to a specific demand system, nor does it necessarily assume a particular type of conduct within the industry. Some potential disadvantages include the fact that GUPPI and UPP may not always correctly predict the sign of the price effect of a merger, and that these measures may either overpredict and undepredict pricing effects; in general this will depend on the nature of competition among non-merging firms, and whether prices are strategic substitutes or strategic complements. Cheung (2011) provides empirical evidence comparing UPP with econometric merger simulation and finds evidence of both type I and type II errors. There is a growing literature that examines the theoretical conditions under which the predictions of UPP and GUPPI accurately predict merger effects, including Carlton (2010), Schmalensee (2009), Willig (2011).

A limitation of the UPP/diversion approach is that it ignores how mergers change the incentives of non-merging firms, as well as the equilibrium quantities. This has led some to consider more complicated (and accurate) alternatives based on pass-through rates (i.e., Jaffe and Weyl (Forthcoming)), with recent empirical work by Miller, Remer, Ryan, and Sheu (2012). This paper remains silent on whether or not Diversion, UPP, and GUPPI accurately capture the price effects of mergers, but instead focuses on how the diversion ratio might be measured by empirical researchers and antitrust practitioners. Farrell and Shapiro (2010) suggest that firms themselves track diversion in their normal course of business, Reynolds and Walters (2008) examine the use of consumer surveys in the UK. Alternatively, diversion could be measured as a prediction of a parametric demand estimation exercise, and Hausman (2010) argues this is the only acceptable method of measuring diversion. We explore an experimentally motivated identification argument for diversion, and then design and conduct a series of field experiments.

Diversion measures the fraction of consumers who switch from product 1 to product 2 as product 1 becomes less attractive (often by raising its price). While it would be impossible (or at least a very bad idea) to randomize whether or not mergers take place, we show that an unintended benefit of the *ceteris parabis* approach embedded in the diversion ratio is that it lends itself to experimental sources of identification. It is relatively straightforward to consider an experiment that increases the price of a single good and measures sales to substitute goods. Another option would be to eliminate product 1 from the choice set and measure substitution directly. By narrowing the focus to the diversion ratio, rather than a full merger simulation, it is now possible to employ a set of tools that have become common in the economics literature on field experiments and treatment effects.

In this paper we design an experiment to measure diversion. However, even though experimental approaches to identification allow treating the diversion ratio as a "treatment effect" of an experiment, we must be careful in considering which treatment effect our experiment measures. We show that the measure of diversion required in UPP analysis represents a marginal treatment effect (MTE), but many experimental designs (including ours) measure an average treatment effect (ATE). We derive an expression for the variance of the diversion ratio, and show that a tradeoff exists between bias and efficiency. For small price increases, an assumption of constant diversion may be reasonable, but the small changes induced in the sales of product 1 may result in quite noisy estimates of diversion. Large changes in p_1 imply a more efficient estimate of diversion, but introduce the possibility for bias if the diversion ratio is not constant. We also derive expressions for diversion under several parametric models of demand, and show how diversion varies over a range of price increases. For example, we show that both the linear demand model and a 'plain vanilla' logit model exhibit constant diversion (treatment effects) for any price increase.

While removing products from consumers' choice sets (or changing prices) may be difficult to do on a national scale, one might be able to measure diversion accurately using smaller, more targeted experiments. In fact, many large retailers such as Target and Wal-Mart frequently engage in experimentation, and online retailers such as Amazon.com and Ebay have automated platforms in place for "A/B-testing." As information technology continues to improve in retail markets, and as firms become more comfortable with experimentation, one could imagine antitrust authorities asking parties to a proposed merger to submit to an experiment executed by an independent third party. One could even imagine both parties entering into an *ex ante* binding agreement that mapped experimental outcomes into a decision on the proposed merger.

An additional challenge in using experiments to measure diversion is that in many retail settings, the overall level of sales is highly variable. In supermarket scanner data, for example, it is not uncommon for category-level sales to vary 30% or more from week to week, while shares of individual products may be quite small ($< 5\%$). This requires either that experiments consider extremely large treatment and control groups, or that adjustments are made for overall heterogeneity. We develop some straightforward adjustments for overall heterogeneity in the context of our experiment.

Our paper considers several hypothetical mergers within the single-serving snack foods industry, and demonstrates how to design and conduct experiments to measure the diversion ratio. We measure diversion by exogenously removing one or two top-selling products from each of three leading manufacturers of snack food products, and observing subsequent substitution patterns. We use a set of sixty vending machine in secure office sites as our experimental "laboratory" for the product removals. We then compare our estimated experimental diversion ratios to those found using structural econometric models, such as a random-coefficients logit model, and a nested logit model. We document cases where these different approaches agree, and where they disagree, and analyze what the likely sources of disagreement are. In general we show that while all approaches generate qualitatively similar predictions, the parametric models tend to under-predict substitution to the very closest substitute, and over-predict substitution to non-substitute products compared to the experimentally-measured treatment effects.

The paper also contributes to a recent discussion about the role of different methods in empirical work going back to Leamer (1983), and discussed recently by Heckman (2010), Leamer (2010), Keane (2010), Sims (2010), Nevo and Whinston (2010), Stock (2010), and Einav and Levin (2010). A central issue in this debate is what role experimental or quasi-experimental methods should play in empirical economic analyses in contrast to structural methods. Angrist and Pischke (2010) complain about the general lack of experimental or quasi-experimental variation in many IO papers. Nevo and Whinston (2010)'s response highlights that most important IO questions, such as prospective merger analysis and counterfactual welfare calculations, are not concerned with measuring treatment effects and do not lend themselves to experimental or quasi-experimental identification strategies. Of course, as screening of potential mergers shifts towards diversion-based measures, this may open the door towards more experimentally-motivated identification strategies.

The paper proceeds as follows. Section 2 lays out a theoretical framework, section 3 describes the snack foods industry, our data, and our experimental design. We describe our calculation of the treatment effect under various models in section 4, and present the results of the analyses in section 5. Section 6 concludes.

2 Theoretical Framework

The first part of this section is purely expositional and does not introduce new results beyond those presented in Farrell and Shapiro (2010) and closely follows Cheung (2011).

For simplicity, consider a single market composed of $f = 1, \ldots, F$ multi-product firms and $j = 1, \ldots, J$ products, where firm f sets the prices of products in set $\mathcal{J}_f$ to maximize profits:

$$\pi_f = \sum_{j \in \mathcal{J}_f} (p_j - c_j) s_j(\mathbf{p}) - C_j$$

Under the assumption of constant marginal costs c_j the FOC for firm f becomes

$$s_j(\mathbf{p}) + \sum_{j \in \mathcal{J}_f} (p_k - c_k) \frac{\partial s_k(\mathbf{p})}{\partial p_j} = 0$$

Let the superscripts (0) and (1) denote pre- and post-merger quantities respectively. Consider the pre- and post-merger FOC of a single-product firm who owns product j and is acquiring product k:

$$\begin{aligned} s_j(\mathbf{p^{(0)}}) + (p_j^{(0)} - c_j) \frac{\partial s_j(\mathbf{p^{(0)}})}{\partial p_j} &= 0 \\ s_j(\mathbf{p^{(1)}}) + (p_j^{(1)} - (1 - e_j) \cdot c_j) \frac{\partial s_j(\mathbf{p^{(1)}})}{\partial p_j} + (p_k^{(1)} - c_k) \frac{\partial s_k(\mathbf{p^{(1)}})}{\partial p_j} &= 0 \end{aligned}$$

Then UPP_j represents the change in the price of $j : p_j^{(1)} - p_j^{(0)}$ that comes about from the difference in the two FOCs, where all other quantities are held fixed at the pre-merger values (i.e.: $\mathbf{p^{(1)}} = \mathbf{p^{(0)}}$ and $p_k^{(1)} = p_k^{(0)}$) and the merger results in cost savings e_j. That is:

$$UPP_j = (p_k^{(0)} - c_k) \cdot \underbrace{\left(\frac{\partial s_j(\mathbf{p^{(0)}})}{\partial p_j}\right)^{-1} \cdot \frac{\partial s_k(\mathbf{p^{(0)}})}{\partial p_j}}_{D_{jk}(\mathbf{p^{(0)}})} - e_j \cdot c_j \tag{1}$$

Here UPP_j measures how the merger affects the opportunity cost of selling an extra unit of j traded off against the marginal cost efficiency. Notice that now the firm internalizes the effect that some fraction of sales of j would have become sales of k. The key input into merger analysis is the quantity $D_{jk}(\mathbf{p^{(0)}})$ or the diversion ratio at the pre-merger prices (and quantities). In words, the diversion ratio measures the fraction of consumers who switch from j to k under pre-merger prices $\mathbf{p^{(0)}}$ when the price of j is increased just enough for a single consumer to cease purchasing j. An alternative to UPP that does not require taking a stance on the marginal cost efficiency e is the Gross Upward Pricing Pressure Index (GUPPI). GUPPI takes the form:

$$GUPPI_j = \frac{p_k^{(0)} - c_k}{p_j^{(0)}} \cdot D_{jk}(\mathbf{p^{(0)}}) \tag{2}$$

The other basic extension we want to consider is when the owner of firm j merges with a firm that controls not only product k, but also product l. In the case where k and l have the same margins $p_k - c_k = p_l - c_l$ then the impact of the three product merger on UPP_j is identical to that in (1) with the exception that it depends on the sum of the diversion ratios $D_{jk} + D_{jl}$.

2.1 Empirically Measuring Diversion

The goal of our paper is to show how the diversion ratio might be estimated experimentally. Part of the rationale given in Farrell and Shapiro (2010) is that the diversion ratio no longer needs to be measured with respect to a particular parametric functional form of demand, and one way to identify D_{jk} might be through conducting an experiment. An obvious experiment would be to exogenously manipulate the price of product j to some random subset of consumers and to measure how the sales of j and k respond. If we let $(\mathbf{p^{(0)}}, \mathbf{p^{(1)}})$ represent the price vectors during the control and treatment, so that $p_l^{(0)} = p_l^{(1)}$, $\forall l \neq j$ and

$p_j^{(1)} = p_j^{(0)} + \Delta p_j$. The diversion ratio can be computed:

$$\widehat{D_{jk}} = \left|\frac{\Delta Q_k}{\Delta Q_j}\right| = \left|\frac{Q_k(\mathbf{p^{(1)}}) - Q_k(\mathbf{p^{(0)}})}{Q_j(\mathbf{p^{(1)}}) - Q_j(\mathbf{p^{(0)}})}\right| = \frac{\int_{p_j^0}^{p_j^1} \frac{\partial q_k(p_j,p_{-j})}{\partial p_j} \partial p_j}{\int_{p_j^0}^{p_j^1} \frac{\partial q_j(p_j,p_{-j})}{\partial p_j} \partial p_j} \tag{3}$$

First we consider the possible bias introduced by considering a larger than infinitesimal increase in price $\Delta p_j \gg 0$. We derive an expression by considering a second order expansion of demand at $\mathbf{p^{(0)}}$:

$$\begin{aligned} q_k(\mathbf{p} + \Delta p_j) &\approx q_k(\mathbf{p}) + \frac{\partial q_k}{\partial p_j}\Delta p_j + \frac{\partial^2 q_k}{\partial p_j^2}(\Delta p_j)^2 + O((\Delta p_j)^3) \\ \frac{q_k(\mathbf{p} + \Delta p_j) - q_k(\mathbf{p})}{\Delta p_j} &\approx \frac{\partial q_k}{\partial p_j} + \frac{\partial^2 q_k}{\partial p_j^2}\Delta p_j + O(\Delta p_j)^2 \end{aligned} \tag{4}$$

This allows us to compute an expression for the bias in D_{jk}:

$$Bias(\widehat{D_{jk}}) \approx -\frac{D_{jk}\frac{\partial^2 q_j}{\partial p_j^2} + \frac{\partial^2 q_k}{\partial p_j^2}}{\frac{\partial q_j}{\partial p_j} + \frac{\partial^2 q_j}{\partial p_j^2}\Delta p_j}\Delta p_j \tag{5}$$

Our expression in (5) shows that the bias of our diversion estimate depends on two things: one is the magnitude of the price increase Δp_j, the second is the curvature of demand $\frac{\partial^2 q_k}{\partial p_j^2}$. This suggests that bias is minimized by experimental designs that consider small price changes. However, small changes in price may lead to noisy measures of Δq. More formally we assume constant diversion $\Delta q_k \approx D_{jk}\Delta q_j$ and construct a measure for the variance of the diversion ratio:

$$Var(\widehat{D_{jk}}) \;\approx\; Var\left(\frac{\Delta q_k}{|\Delta q_j|}\right) \approx \frac{1}{\Delta q_j^2}\left(D_{jk}^2\sigma_{\Delta q_j}^2 + \sigma_{\Delta q_k}^2 - 2D_{jk}\rho\sigma_{\Delta q_j}\sigma_{\Delta q_k}\right) \tag{6}$$

This implies a bias-variance tradeoff when estimating diversion. If our primary concern is that the curvature of demand is steep, it suggests considering a small price increase. However, if our primary concern is that sales are highly variable, we should consider a larger price increase.

Instead of viewing our experiment as a biased measure of the marginal treatment effect, we can instead ask: "What quantity does our experimental diversion measure estimate?"

We can derive an approximate relationship for our experimental diversion measure ED_{jk}:

$$\widehat{ED_{jk}} = \frac{1}{\Delta q_j} \int_{p_j^0}^{p_j^0+\Delta p_1} \underbrace{\frac{\partial q_k}{\partial q_j}}_{D_{jk}(\mathbf{p})} \left| \frac{\partial q_j}{\partial p_j} \right| \partial p_j \tag{7}$$

Thus an experiment varying the price p_j measures the weighted average of diversion ratios for different price increases, where the weights correspond to the lost sales of j at a particular p_j as a fraction of all lost sales of j.

The experiment in our empirical example does not employ actual price changes, instead the focal product is removed from the consumer's choice set. This is equivalent to increasing the price p_j to the choke price where $q_j(p_j^c, \mathbf{p}_{-\mathbf{j}}^{(\mathbf{0})}) = 0$. This has the advantage that it minimizes the variance expression in (6). The other derivations in this section provide insight into instances in which a product removal provides an accurate measure of diversion: (a) when the curvature of demand is small ($\frac{\partial^2 q_k}{\partial p_j^2} \approx 0$), (b) when the true diversion ratio is constant (or nearly constant) $D_{jk}(\mathbf{p}) = D_{jk}$, or (c) when demand for j is steepest near the market price $\left|\frac{\partial q_j(p_j, \mathbf{p}_{-\mathbf{j}}^{(\mathbf{0})})}{\partial p_j}\right| \gg \left|\frac{\partial q_j(p_j+\delta, \mathbf{p}_{-\mathbf{j}}^{(\mathbf{0})})}{\partial p_j}\right|$.

In our example, it might seem reasonable that customers who substitute away from a *Snickers* bar after a five cent price increase switch to *Reese's Peanut Butter Cup* at the same rate as after a 25 cent price increase, where the only difference is the number of overall consumers leaving *Snickers*. However in a different industry, this may no longer seem as reasonable. For example, we might expect buyers of a Toyota Prius to substitute primarily to other cheap, fuel-efficient cars when faced with a small price increase (from the market price of $25,000 to $25,500), but we might expect some substitution to luxury cars when facing a larger price increase (to $50,000). If demand (in units) for the Prius falls rapidly with a small price increase, so that residual demand (and the potential impact of further price increases) is small, then an experiment that removes it from the choice set may provide an accurate measure of diversion to remaining products, even though the diversion ratio itself is not constant.

In the Appendix we derive explicit relationships for the diversion ratio for several well known parametric demand functions. There are two functional forms for demand that exhibit constant diversion and are always unbiased: the first is linear demand, for which $\frac{\partial^2 q_k}{\partial p_j^2} = 0$, $\forall j, k$. The second is the IIA logit model, for which $D_{jk} = -\frac{\partial^2 q_k}{\partial p_j^2} / \frac{\partial^2 q_j}{\partial p_j^2}$. We derive these relationships, as well as expressions for diversion under other demand models in the Appendix, and show that random coefficients logit demand, and CES demands (including log-linear demand) do not generally exhibit constant diversion.

3 Description of Data and Industry

Globally, the snack foods industry is a $300 billion a year business, composed of a number of large, well-known firms and some of the most heavily advertised global brands. Mars Incorporated reported over $50 billion in revenue in 2010, and represents the third largest privately-held firm in the US. Other substantial players include Hershey, Nestle, Kraft, Kellogg, and the Frito-Lay division of PepsiCo. While the snack-food industry as a whole might not appear highly concentrated, sales within product categories can be very concentrated. For example, Frito-Lay comprises around 40% of all savory snack sales in the United States, and reported over $13 billion in US revenues last year, but its sales outside the salty-snack category are minimal, coming mostly through parent PepsiCo's Quaker Oats brand and the sales of *Quaker Chewy Granola Bars*.[2] We report HHI's at both the category level and for all vending products in Table 1 from the region of the U.S. that includes our vending operator. If the relevant market is defined at the category level, all categories are considered highly concentrated, with HHIs in the range of roughly 4500-6300. At the overall vending machine level, the HHI is below the critical threshold of 2500. Any evaluation of a merger in this industry would hinge on the closeness of competition, and thus require measuring diversion.

Over the last 25 years, the industry has been characterized by a large amount of merger and acquisition activity, both on the level of individual brands and entire firms. For example, the *Famous Amos* cookie brand has been owned by the Kellogg Company since 2001. Between 1985 and 2001, Famous Amos cookies were owned by at least seven firms, including the Keebler Cookie Company (acquired by Kellogg in 2001), Presidential Baking Company (acquired by Keebler in 1998), as well as by other snack food makers and private equity firms. *Zoo Animal Crackers* have a similarly complicated history owned by cracker manufacturer Austin Quality Foods, before they too were acquired by the Keebler Cookie Co. (who in turn was acquired by Kellogg).[3]

[2] Most analysts believe Pepsi's acquisition of Quaker Oats in 2001 was unrelated to its namesake business but rather for Quaker Oats' ownership of Gatorade, a close competitor in the soft drink business.

[3] A landmark case in market definition was brought by *Tastykake* in attempt to block the acquisition of *Drake* (the maker of Ring-Dings) by *Ralston-Purina's Hostess* brand (the maker of Twinkies). That case established the importance of geographically significant markets, as Drake's had only 2% marketshare nationwide but a much larger share in the Northeast (including 50% of the New York market). Tastykake also successfully argued that the relevant market was single-serving snack cakes rather than a broad category of snack foods involving cookies, candy bars. [Tasty Baking Co. v. Ralston Purina, Inc., 653 F. Supp. 1250

Our study measures diversion through the lens of a single medium-sized retail vending operator in the Chicago metropolitan area, MarkVend. Each of MarkVend's machines internally records price and quantity information. The data track total vends and revenues since the last service visit on an item-level basis (but do not include time-stamps for each sale). Any given machine can carry roughly 35 products at one time, depending on configuration.

We observe retail and wholesale prices for each product at each service visit during our 38-month panel. There is relatively little price variation within a site, and almost no price variation within a category (e.g., chocolate candy) at a site. This is helpful from an experimental design perspective, but can pose a challenge to demand estimation. Very few "natural" stock-outs occur at our set of machines.[4] Most changes to the set of products available to consumers are a result of product rotations, new product introductions, and product retirements. Over all sites and months, we observe 185 unique products. We consolidate some products with very low levels of sales using similar products within a category produced by the same manufacturer, until we are left with the 73 'products' that form the basis of the rest of our exercise.[5]

In addition to the data from Mark Vend, we also collect data on the characteristics of each product online and through industry trade sources.[6] For each product, we note its manufacturer, as well as the following set of product characteristics: package size, number of servings, and nutritional information.[7]

3.1 Experimental Design

We ran four experimental treatments with the help of the Mark Vend Company. These represent a subset of a larger group of experiments we have used in other projects, such as Conlon and Mortimer (2010) and Conlon and Mortimer (2013a). Our experiment follows 60 snack machines located in professional office buildings and serviced by MarkVend. Most of the customers at these sites are 'white-collar' employees of law firms and insurance companies. Our goal in selecting the machines was to choose machines that could be analyzed together, in

- Dist. Court, ED Pennsylvania 1987]

[4]Mark Vend commits to a low level of stock-out events in its service contracts.

[5]For example, we combine Milky Way Midnight with Milky Way, and Ruffles Original with Ruffles Sour Cream and Cheddar.

[6]For consolidated products, we collect data on product characteristics at the disaggregated level. The characteristics of the consolidated product are computed as the weighted average of the characteristics of the component products, using vends to weight. In many cases, the observable characteristics are identical.

[7]Nutritional information includes weight, calories, fat calories, sodium, fiber, sugars, protein, carbohydrates, and cholesterol.

order to be able to run each experiment over a shorter period of time across more machines.[8] These machines were also located on routes that were staffed by experienced drivers, which maximized the chance that the experiment would be successfully implemented. The 60 machines used for each experiment were distributed across five of Mark Vend's clients, which had between 3 and 21 machines each. The largest client had two sets of floors serviced on different days, and we divided this client into two sites. Generally, each site is spread across multiple floors in a single high-rise office building, with machines located on each floor.

For each treatment, we removed a product from all machines at a client site for a period of 2.5 to 3 weeks. We conducted eight treatments, four of which are analyzed here. In two of the four treatments we simultaneously removed the two best-selling products from either (a) chocolate maker Mars Incorporated (Snickers and Peanut M&Ms), or (b) salty-snack producer PepsiCo (Doritos Nacho Cheese and Cheetos Crunchy). For the third and fourth treatments we individually removed two products owned by Kellogg's: Famous Amos Chocolate Chip Cookies, and Zoo Animal Crackers. We chose to run these last two treatments separately in part because Animal Crackers are difficult to categorize, and close substitutes are less obvious.[9]

Whenever a product was experimentally stocked-out, poster-card announcements were placed at the front of the empty product column. The announcements read "This product is temporarily unavailable. We apologize for any inconvenience." The purpose of the card was two-fold: first, we wanted to avoid dynamic effects on sales as much as possible, and second, the firm wanted to minimize the number of phone calls received in response to the stock-out events. The dates of the interventions range from June 2007 to September 2008, with all removals run during the months of May - October. We collected data for all machines for just over three years, from January of 2006 until February of 2009. During each 2-3 week experimental period, most machines receive service visits about three times. However, the length of service visits varies across machines, with some machines visited more frequently than others.

The cost of the experiment consisted primarily of driver costs. Drivers had to spend extra time removing and reintroducing products to machines, and the driver dispatcher had to spend time instructing the drivers, tracking the dates of each experiment, and reviewing

[8]Many high-volume machines are located in public areas (e.g., museums or hospitals), and have demand (and populations) that varies enormously from one day to the next, so we did not use machines of this nature. In contrast, the work-force populations at our experimental sites are relatively homogenous.

[9]The remaining four treatments involved the individual removal of each individual chocolate or salty-snack product from our first two treatments. These are analyzed in Conlon and Mortimer (2010) and Conlon and Mortimer (2013a), but are less relevant for this paper.

the data as they were collected. Drivers are generally paid a small commission on the sales on their routes, so if sales levels fell dramatically as a result of the experiments, their commissions could be affected. Tracking commissions and extra minutes on each route for each driver would have been prohibitively expensive to do, and so drivers were provided with \$25 gift cards for gasoline during each week in which a product was removed on their route to compensate them for the extra time and the potential for lower commissions.

Our experiment differs somewhat from an ideal experiment. Ideally, we would be able to randomize the choice set on an individual level, though technologically that is difficult in both vending and traditional brick and mortar contexts. In contrast, online retailers are capable of showing consumers different sets of products and prices simultaneously. This leaves our design susceptible to contamination if for example, Kraft runs a large advertising campaign for Planters Peanuts that corresponds to the timing of one of our experiments. Additionally, because we remove all of the products at an entire client site for a period of 2.5 to 3 weeks, we lack a contemporaneous group of untreated machines. We chose this design, rather than randomly staggering the product removals, because we (and the participating firms) were afraid consumers might travel from floor to floor searching for stocked out products. This design consideration prevents us from using control machines in the same building, and makes it more difficult to capture weekly variation in sales due to unrelated factors, such as a law firm taking a big case to trial, or accountants during quarterly reporting season.[10] Despite the imperfections of field experiments in general, these are often the kinds of experiments run by firms in their regular course of business, and may most closely approximate the type of experimental information that a firm may already have available at the time when a proposed merger is initially screened.

4 Analyses of the Experimental Outcomes

The goal of our experiment is to determine how sales are diverted away from best-selling products. In principle, this calculation is straightforward. In practice, however, there are two challenges in implementing the experiment and interpreting the data generated by it. The first challenge is that there is a large amount of variation in overall sales at the weekly level independent of our product removals. This weekly variation in overall sales is common in retail environments. We often observe week-over-week sales that vary by over 20%, with

[10] On balance, we thought that people traveling from floor to floor was a larger concern. It also has the additional benefit that we can aggregate over all machines at a client site, and treat the entire site as if it were a single machine.

no single product having more than 4.5% market share. This can be seen in Figure 1, which plots the overall sales of all machines from one of the sites in our sample on a weekly basis. For example, a law firm may have a large case going to trial in a given month, and vend levels will increase at the firm during that period. In our particular setting, many of the treatments were run during the summer of 2007, which was a high-point in demand at these sites, most likely due to macroeconomic conditions. In this case, using a simple control like previous weeks' sales, or overall average sales, could result in unreasonable treatment effects, such as sales increasing due to product removals, or sales decreasing by more than the sales of the removed products. This requires either adjustments to control for aggregate volatility in demand, or an alternative design that considers a substantially larger treatment group.[11]

The second challenge is that the data are recorded at the level of a service visit to a vending machine. It is sometimes more convenient to organize observations by week, rather than by visit, because different visits occur on different days of the week. In order to do this, we assign sales to a uniform measure of a week, assuming that sales are distributed uniformly among the business days in a service interval. We allow our definition of when weeks start and end to depend on the client site and experiment because different experimental treatments start on different days of the week.[12]

4.1 Computing Treatment Effects

In order to compute treatment effects, we treat all machines on all floors of a particular office building as a single machine. We do this for two reasons, the first is that sales at the individual machine level are quite small and can vary quite a bit over time. This would make it hard to measure any kind of effect. The second is that we don't want to worry about consumers going from machine to machine searching for missing products, or additional noise in demand created by a long meeting on a particular floor, etc. We address the problem of aggregate volatility by choosing the set of relevant control weeks S_t with two restrictions born out of consumer theory: (1) removing products cannot increase total sales, and (2) removing products cannot reduce total sales by more than the sales of the removed products. Let q_{jt} denote the sales of product j in site-week t, and superscript 1 denote sales when a treated product(s) is removed, and superscript 0 denote sales when a treated product(s) is available. If Q_t denotes aggregate sales during client-week t, and q^0_{js} denotes the sales of the removed product during control week s then these two restrictions imply that the set of control weeks

[11]However, it is worth pointing out that our dataset already consists of nearly 3 million individual purchases, and likely mimics the nature and scale of experiments run by firms in their natural course of business.

[12]At some site-experiment pairs, weeks run Tuesday to Monday, while others run Thursday to Wednesday.

is given by:

$$\{s : s \neq t, Q_s^0 - Q_t^1 \in [0, q_{js}^0]\} \tag{8}$$

Notice that this is the same as placing a restriction on diversion from the removed product j to all other products, which is:

$$\sum_{k \neq j} D_{jk} \in [0, 100\%] \tag{9}$$

The problem with a direct implementation of (8) is that weeks with high sales of the focal product q_{js}^0 are more likely to be included in the control. This selection bias would understate the diversion ratio. We propose a slight modification of (8) which is unbiased. We replace q_{js}^0 with $\widehat{q_{js}^0} = E[q_{js}^0 | Q_s^0]$. An easy way to obtain the expectation is to run an OLS regression of q_{js}^0 on Q_s^0. We run one regression for each client-site and report the results for one of the client-sites in Table 2. We use our definition of control weeks s to compute the expected control sales that correspond to treatment week t as:

$$S_t = \{s : s \neq t, Q_s^0 - Q_t^1 \in [0, \hat{b_0} + \hat{b_1} Q_s^0]\} \tag{10}$$

In 528 of 634 machine-treatment weeks, at least one corresponding control week is found for each treatment observation. On average we find 24 control observations per treatment observation.[13]

For each treatment week t we can compute the treatment effect and diversion ratio as:

$$\widehat{D_{jk}^{diff}} = \gamma \frac{\Delta q_{kt}}{\Delta q_{jt}} \tag{11}$$

$$\Delta q_{kt} = q_{kt}^1 - \frac{AV_k^{(1)}}{AV_k^{(0)}} \cdot \frac{1}{\# S_t} \sum_{s \in S_t} q_{ks}^0 \tag{12}$$

where S_t is the set of control weeks that corresponds to a particular treatment week t. Ignoring the $\frac{AV_k^{(1)}}{AV_k^{(0)}}$ term, this is just the difference between sales during the treatment weeks and average sales during the corresponding control weeks.

The role of the $\frac{AV_k^{(1)}}{AV_k^{(0)}}$ term is to adjust for changes in the product availability between the treatment and control period. It has the effect of transforming the product availability

[13] If we were focused on a particular merger, we could also condition S_t on a specific set of relevant products in which we were interested, for example, the products of a close competitor.

to match the product availability during the treatment period. Thus we can consider (11) as measuring the average treatment on the treated (ATT). The units for AV_k correspond to the aggregate sales in the weeks where k was available.

We report a subset of the availability adjustments in Table 4 for the Mars (M&M Peanut and Snickers) product removals.[14] While this adjustment addresses changes in availability of substitute products, it no longer gets the aggregate diversion patterns correct. Even with our restriction on control weeks in (10), we can generate diversion ratios that are less than zero or greater than 100% in aggregate after the availability adjustments. Therefore, we rescale the adjusted change in vends for products with positive diversion by a constant (treatment-specific) factor, γ, for all products so that the aggregate diversion ratio is the same before and after our availability adjustment. This is demonstrated in Table 3.

Table 4 shows the availability adjustments for individual products for a single treatment (i.e., the Snickers-Peanut M&M removal). We note that the adjustments are generally quite small, as availability is roughly the same during the treatment and control weeks. However, in a few cases the adjustment is crucial. For example, *Reese's Peanut Butter Cups* is a close competitor to the two Mars products *Snickers* and *Peanut M&Ms*, and sees the largest change in unadjusted sales (174.2 vs 56.5, or 118.2 units). After the adjustment, the change in sales is considerably smaller (33.4 units instead of 118.2 units) but still quite substantial. Similarly, *Salty Other* shows a negative change in sales as a result of the Mars product removals, implying that it is a complement rather than a substitute, but after adjusting for declining availability, the sign changes.

Once we have constructed our restricted set of treatment weeks and the set of control weeks that corresponds to each, inference is fairly straightforward. We aggregate to the manufacturer level, and use (11) to construct a set of pseudo-observations, and employ a t-test for the difference.

4.2 Regression-based Approach

An obvious approach to measuring the effects of the experiment would be to consider a regression that regresses sales of product j in week t and machine m on a treatment indicator, and estimate a different coefficient for each product. This would not address the primary challenge that aggregate sales vary dramatically at the weekly level. Week-specific (or week-client-specific) fixed effects both explain little of the variation at the machine-week level and

[14] The full set of adjustments for all product and all experiments are available from the authors upon request.

have the problem that they are collinear with the treatment. To address this problem we instead use manufacturer-specific marketshares rather than sales as our dependent variable:[15]

$$s_{fmt} \equiv \frac{q_{fmt}}{M_{mt}} = \alpha_{fm} + \beta_f \times treatment_{mt} + \varepsilon_{fmt} \tag{13}$$

This requires a choice of the market size M_{mt}. One option would be to use the sales of all products Q_{mt}, but that has the problem that it would not accurately capture substitution to the outside good. Instead, we let $M_{mt} = M_m$ be the maximum weekly sales for that particular machine.[16] This specification addresses the cross-machine variation in tastes for the products of different manufacturers. Though we do not include them in our primary specification, we could also include additional regressors for the availability of other competing products.

Under this specification, we construct the expected change in sales of k, and the corresponding diversion ratio implied by the experiment:

$$\widehat{D_{jf}} = \frac{E[\Delta q_f]}{E[\Delta q_j]} = \gamma \frac{\widehat{\beta_f} \cdot M_f}{\widehat{\beta_j} \cdot M_j} \tag{14}$$

where M_j, and M_f denote the availability (in market size terms) of the removed product(s), and the substitute manufacturer, respectively. Appropriate choices of M_j, and M_f allow us to compute the average treatment on the treated (ATT), average treatment on control (ATC), or the average treatment effect (ATE). If all products were always available, all three of these quantities would always be the same. In practice these are easy to compute as they involve predicting the change in quantity for each observation and then aggregating across the appropriate set of weeks.

As an example, consider the case of single product firms where the focal product is always available (we consider the focal product to be "available" but treated during our treatment period). If the product was available for 90% of treatment consumers, and 10% of control consumers we would adjust the ratio of $\widehat{\beta}$ coefficients by 0.9 and 0.1 respectively to compute the ATT and ATC. If the sample were evenly split between treatment and control, then the ATE would imply rescaling diversion by 0.5.

Just like in the differences case, this approach gets the relative diversion ratios correct,

[15] As margins do not vary at the manufacturer-category level, and UPP depends only on manufacturer-level diversion, we aggregate sales to the level of a manufacturer f rather than a product j, $q_{fmt} = \sum_{j \in \mathcal{J}_F} q_{jmt}$.

[16] As a robustness test we've experimented with different specifications that also use common shocks across weeks. Other than some obvious holiday weeks, the weekly variation tends to be uncorrelated across machines. About 65% of the variation in weekly machine sales is explained by the machine, and only about 10% is explained by the week.

but once we aggregate across weeks with different availability, it does not get the aggregate scale of diversion ratios correct. We address this by rescaling where γ is chosen to match the aggregate diversion (for all manufacturers with positive diversion) from our unadjusted differences.

4.3 Parametric Specifications

In addition to computing treatment effects, we also specify two parametric models of demand: nested logit and random-coefficients logit, which are estimated from the full dataset (including weeks of observational data that do not meet any of our control criteria).

We consider a model of utility where consumer i receives utility from choosing product j in market t of:

$$u_{ijt} = \delta_j + \xi_t + \mu_{ijt} + \varepsilon_{ijt} \tag{15}$$

The parameter δ_j is a product-specific intercept that captures the mean utility of product j, ξ_t captures market level heterogeneity in demand, and μ_{ijt} captures individual-specific correlation in tastes for products.

In the case where $(\mu_{ijt} + \varepsilon_{ijt})$ is distributed generalized extreme value, the error terms allow for correlation among products within a pre-specified group, but otherwise assume no correlation. This produces the well-known nested-logit model of McFadden (1978). In this model consumers first choose a product category l composed of products g_l, and then choose a specific product j within that group. The resulting choice probability for product j in market t is given by the closed-form expression:

$$p_{jt}(\delta, \lambda, a_t) \quad = \quad \frac{e^{(d_j+\xi_t)/\lambda_l}\left(\sum_{k\in g_l\cap a_t} e^{(\delta_k+\xi_t)/\lambda_l}\right)^{\lambda_l-1}}{\sum_{\forall l}\left(\sum_{k\in g_l\cap a_t} e^{(\delta_k+\xi_t)/\lambda_l}\right)^{\lambda_l}} \tag{16}$$

where the parameter λ_l governs within-group correlation, and a_t is the set of available products in market t.[17]

The random-coefficients logit allows for correlation in tastes across observed product characteristics. This correlation in tastes is captured by allowing the term μ_{ijt} to be distributed

[17] Note that this is not the IV regression/'within-group share' presentation of the nested-logit model in Berry (1994), in which σ provides a measure of the correlation of choices within a nest. Roughly speaking, in the notation used here, $\lambda = 1$ corresponds to the plain logit, and $(1 - \lambda)$ provides a measure of the 'correlation' of choices within a nest (as in McFadden (1978)). The parameter λ is sometimes referred to as the 'dissimiliarity parameter.'

according to $f(\mu_{ijt}|\theta)$. A common specification is to allow consumers to have independent normally distributed tastes for product characteristics, so that $\mu_{ijt} = \sum_l \sigma_l \nu_{ilt} x_{jl}$ where $\nu_{ilt} \sim N(0,1)$ and σ_l represents the standard deviation of the heterogeneous taste for product characteristic x_{jl}. The resulting choice probabilities are a mixture over the logit choice probabilities for many different values of μ_{ijt}, shown here:

$$p_{jt}(\delta, \theta, a_t) = \int \frac{e^{d_j+\xi_t+\sum_l \sigma_l \nu_{ilt} x_{jl}}}{1+\sum_{k\in a_t} e^{d_k+\xi_t+\sum_l \sigma_l \nu_{ilt} x_{kl}}} f(\nu) d\,\nu \tag{17}$$

Estimation then proceeds by full information maximum likelihood (FIML) in the case of nested logit or maximum simulated likelihood (MSL) in the case of the random coefficients. The log-likelihood is:

$$\widehat{\theta}_{MLE} = \arg\max_{\theta} \sum_{j,t} q_{j,t} \log(p_{jt}(\delta, \theta; a_t))$$

In both the nested-logit and random-coefficient models we include an additional ξ_t for either a machine-visit pair (15,256 fixed effects), or a choice set (2,710 fixed effects). For the nested-logit model, we allow for heterogeneous tastes across five major product categories or nests: chocolate candy, non-chocolate candy, cookie, salty snack, and other.[18] For the random-coefficients specification, we allow for three random coefficients, corresponding to consumer tastes for salt, sugar, and nut content.[19]

In Table 5, we report the parameter estimates for our preferred specification of the random-coefficients and nested-logit models. We considered a large number of alternative models, but presented the results selected by the Bayesian Information Crtieria (BIC). All of our estimates included 73 product-specific intercepts. In the case of the nested logit model, this included 5 nests (Chocolate, Non-Chocolate Candy, Cookie/Pastry, SaltySnack, and Other). For the random coefficients model, this included only three normally distributed

[18]The vending operator defines categories in the same way. "Other" includes products such as peanuts, fruit snacks, crackers, and granola bars.

[19]We do not allow for a random coefficient on price because of the relative lack of price variation in the vending machines. We also do not include random coefficients on any discrete variables (such as whether or not a product contains chocolate). As we discuss in Conlon and Mortimer (2013b), the lack of variation in a continuous variable (e.g., price) implies that random coefficients on categorical variables may not be identified when product dummies are included in estimation. We did estimate a number of alternative specifications in which we include random coefficients on other continuous variables, such as carbohydrates, fat, or calories. In general, the additional parameters were not significantly different from zero, and they had no appreciable effect on the results of any prediction exercises.

random coefficients (Salt, Sugar, and Nut Content), though we considered additional specifications including coefficients on Fat, Calories, Chocolate, Protein, Cheese, and other product characteristics. Additionally we estimated both sets of models with both choice set fixed demand shifters ξ_t and machine-visit level demand shifters. While the extra demand shifters improved the fit, BIC preferred the smaller set of choice-set fixed effects.

Computing the diversion ratios is a straightforward prediction from the model:

$$\widehat{D_{jk}}(a_t, \xi_t, \hat{\theta}) = \frac{p_k(a_t \setminus j, \xi_t, \hat{\theta}) - p_k(a_t, \xi_t, \hat{\theta})}{p_j(a_t, \xi_t, \hat{\theta})}$$

The important decision is how to determine the choice set used in the counterfactual prediction exercise. We use a single choice set based on the most commonly available products during our treatment period (so as to approximate ATT). An alternative would be to consider a representative choice set based on some other period, or to predict the change in vends for each market observed in the data and then aggregate to obtain a choice-set weighted diversion prediction (similar to how we construct the OLS prediction). We choose the simpler representative choice set, because we believe it better approximates the approach likely to be taken by empirical researchers. We also choose a value of ξ_t that corresponds to the treatment period.

4.4 Identification and Parameter Estimates

The treatment effects approach and the parametric model rely on two different sources of identification. Formal nonparametric identification results for random utility models such as Berry and Haile (2010) or Fox and Gandhi (2010) often rely on variation across markets in continuous characteristics such as price. This is unavailable in the vending setting, since there is little to no price variation. Instead, the parametric models are identified through discrete changes in the choice set, primarily through product rotations. The intuition is that *Snickers* and *Milky Way* may look similar in terms of observable characteristics (a *Snickers* is essentially a *Milky Way* with peanuts), but have different marketshares. In fact, *Snickers* often outsells *Milky Way* 3:1 or better, which leads us to conclude that Snickers offers higher mean utility to consumers. At the same time, sales may respond differently to the availability of a third product. For example, if *Planters Peanuts* is introduced to the choice set, and it reduces sales of *Snickers* relatively more than *Milky Way*, we might conclude there are heterogeneous preferences for peanuts. One challenge of this approach, is that while we might observe many differences in relative substitution patterns across products, we must in essence project them onto a lower dimensional basis of random coefficients. Thus if our model did not include a random parameter for peanuts, we would have to explain those tastes with something else, like “salt.”

In the absence of experimental variation, many of the best-selling products are essentially always stocked by the retailer. Therefore we learn about the closeness of competition between popular products by observing how market shares respond to the availability of a third (often much less popular) product. The identifying variation comes through the fact that we observe 2,710 different choice sets in 15,256 service visits. If, for example, all machines stocked exactly the same set of products every week, we would only have a single choice set, and would struggle to identify any nonlinear parameters.[20]

While the product rotations are crucial to the identification of the parametric models, they are somewhat of a nuisance to the identification of the treatment effects model. Product rotations introduce additional heterogeneity for which we must control, or we risk introducing bias into the estimated treatment effect. The ideal identification setting for the treatment effect would be a case with no non-experimental variation in either prices or the set of available products. Thus the treatment effect estimator should perform well precisely when the parametric demand model may be poorly identified, and vice versa. This creates an inherent problem in any setting where we want to evaluate the relative performance of the two approaches. In contrast, both the treatment effects approach and the discrete choice models benefit from experimental variation in the choice set.

5 Results

Our primary goal has been both to show how to measure diversion ratios experimentally, but also to understand how those experimentally measured diversion ratios compare to diversion ratios obtained from common parametric models of demand.

Table 6 shows the diversion computed for the top 5 substitutes under the treatment effects approach and the diversion computed under random coefficients, and nested logit models for each experiment. There is a general pattern that emerges. The logit-type models and the treatment effects approach predict a similar order for substitutes (i.e.: the best substitute, the second best substitute, and so on). Also, they tend to predict similar diversion ratios for the third, fourth and fifth best substitute. For example in the Snickers and Peanut M&M removal, the treatment effects approach and the random coefficients logit approach predict around 4% diversion to *Reese's Peanut Butter Cups* and *M&M Milk Chocolate*. However,

[20]Typically this problem could be alleviated by seeing variation in prices within a product over time.

there tends to be a large discrepancy for diversion to the best-substitute product in every removal. The treatment effect predicts nearly 14.2% of consumers substituting to Twix, while the random coefficients model predicts only 6.4%. This effect is large enough that it could lead to incorrect conclusions about upward pricing pressure or a potential merger.

It is important to point out that our results should not be interpreted as showing that the treatment effects approach is always preferred to the parametric demand estimation approach. In the *Famous Amos Cookie* removal, the treatment effect shows that 20.7% of consumers switch from *Famous Amos Chocolate Chip Cookies* to *Sun Chips*, while the random coefficient model predicts diversion $< 1\%$ (the products are quite dissimilar). While this effect appears large in the raw data, it represents an effect that most industry experts would not likely believe. Moreover, the effect is not precisely estimated as indicated by table 7. In part, this may be because we have imperfectly adjusted for the increased availability of *Sun Chips*, or because we have failed to adjust for some close competitor of *Sun Chips*, or that overall sales levels of *Sun Chips* are highly variable.[21] While the parametric demand models observe the reduced availability of *Smartfood Popcorn* and use that to improve its estimates of parameters, the treatment effects model is now faced with confounding variation.

Table 7 shows diversion ratios aggregated to the level of a manufacturer. The results exhibit the same pattern as before, in which the treatment effects predict much more substitution to the closest substitutes than the random coefficients and nested logit models do. For merger analysis, correctly predicting diversion to one or two closest substitutes might be the most important feature of a model or identification strategy, because those are the potential mergers about which we should be most concerned. The other potential anomaly is that after adjusting for product availability, the aggregate diversion measures from the paired differences approach can be unrealistically high ($> 240\%$). In the columns labeled "Adj. Differences" and "Adj. OLS," we re-scale all of the diversion estimates so that the aggregate diversion (among manufacturers with positive diversion) matches the original aggregate diversion in the paired (but unadjusted for availability) data. Much of the differences between the parametric models and the treatment effects approach can be attributed to substitution to the outside good. The parametric models predict that consumers will be diverted to some other product at a rate of approximately 30-45%, while the paired and OLS approaches predict that consumers will be diverted at a rate of 65-75%. All of the models broadly

[21] Imagine that our treatment weeks at our largest location are correlated with reduced availability of *Smartfood Popcorn* for reasons unrelated to our removal, such as supply problems at the warehouse. (It actually is!) In that case we could think about the actual treatment as removing *Famous Amos Cookies* and *Smartfood Popcorn* simultaneously.

predict the same relative magnitudes, but the parametric models predict smaller aggregate magnitudes. A different parametrization of the outside share might lead to more similar estimates.

In table 8 we compute the gross upward pricing pressure index (GUPPI), which is $GUPPI_j = D_{jk}\frac{p_k - c_k}{p_j}$. We exploit the fact that we observe the wholesale contracts between the retailer and the manufacturer, and let the wholesale prices serve as p_j, and p_k in the above expression. We assume that all products have a manufacturing cost of \$0.15, and exploit the fact that within manufacturer-category there is no wholesale price variation: (*Snickers* and *3 Musketeers* have the same wholesale price even though one is much more popular than the other). Under a symmetry assumption (not true in our example) the critical value for GUPPI is generally 10%, because it corresponds to a 'small but significant and non-transitory price increase,' or SSNIP, of 5% under the hypothetical monopolist test. If we apply the same 10% threshold to our results we find that a Mars-Hershey merger would likely attract further scrutiny, but not a Mars-Nestle merger (at least not for the wholesale prices of Snickers and Peanut M&Ms). Meanwhile, the acquisition of the Mars or Pepsi product portfolio might place upward pricing pressure on Zoo Animal Crackers, as they exhibit large point estimates for GUPPI under the treatment effects approach, though diversion is not statistically significant from zero, and the effects are imprecise. Similarly, the Famous Amos cookies show a large but imprecise point estimate for the acquisition of the Pepsi portfolio (driven by an inexplicable rise in the sales of Sun Chips).

6 Conclusion

Under the revised 2010 Horizontal Merger Guidelines, the focus on the diversion ratio and the "single product merger simulation" approach in UPP and GUPPI where all other prices and quantities are held fixed, imply that the diversion ratio serves as a single sufficient statistic for at least the initial screening stage of merger evaluation. We show that the diversion ratio has the attractive empirical property that it can be estimated via experimental or quasi-experimental techniques in a relatively straightforward manner. Our hope is that this makes a well-developed set of tools available both to researchers in Industrial Organization, and also to antitrust practitioners.

At the same time, we are quick to point out that while the diversion ratio can be obtained experimentally, it is not trivial, and researchers should think carefully about which treatment effect their experiment (or quasi-experiment) is actually identifying; as well as what identifying assumptions required for estimating the diversion ratio implicitly assume about the structure of demand. An important characteristic of many retail settings is that even category level sales are much more variable than most product level market shares. We partially alleviate this problem by employing full removal of the product, rather than marginal price increases, and derive conditions under which this may provide a reasonable approximation to the marginal diversion ratio. Even under our design, the overall variability in the sales levels can make diversion ratios difficult to measure. If firms submitted to similar experiments as part of a merger review process, or made previous experimental results available, they would likely focus on a narrower set of potential products than we considered in our setting.

References

Angrist, J., and J.-S. Pischke (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *The Journal of Economic Perspectives*, 24(2), 3–30.

Berry, S. (1994): "Estimating discrete-choice models of product differentiation," *RAND Journal of Economics*, 25(2), 242–261.

Berry, S., and P. Haile (2010): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Working Paper.

Carlton, D. W. (2010): "Revising the Horizontal Merger Guidelines," *Journal of Competition Law and Economics*, 6(3), 619–652.

Chetty, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1(1), 451–488.

Cheung, L. (2011): "The Upward Pricing Pressure Test for Merger Analysis: An Empirical Examination," Working Paper.

Conlon, C., and J. H. Mortimer (2010): "Effects of Product Availability: Experimental Evidence," NBER Working Paper 16506.

——— (2013a): "All Units Discount: Experimental Evidence from the Vending Industry," Working Paper.

——— (2013b): "Demand Estimation Under Incomplete Product Availability," *American Economic Journal: Microeconomics*, 5(4), 1–30.

Einav, L., and J. Levin (2010): "Empirical Industrial Organization: A Progress Report," *The Journal of Economic Perspectives*, 24(2), 145–162.

Farrell, J., and C. Shapiro (2010): "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition," *The B.E. Journal of Theoretical Economics*, 10(1), 1–41.

Fox, J., and A. Gandhi (2010): "Nonparametric Identification and Estimation of Random Coefficients in Nonlinear Economic Models," Working Paper.

Hausman, J. A. (2010): "2010 Merger Guidelines: Empirical Analysis," Working Paper.

Heckman, J. J. (2010): "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature*, 48(2), 356–398.

Jaffe, S., and E. G. Weyl (Forthcoming): "The First Order Approach to Merger Analysis," *American Economic Journal: Microeconomics*.

Keane, M. (2010): "A Structural Perspective on the Experimentalist School," *The Journal of Economic Perspectives*, 24(2), 47–58.

Leamer, E. (1983): "Let's Take the Con Out of Econometrics," *American Economic Review*, 75(3), 308–313.

——— (2010): "Tantalus on the Road to Asymptopia," *The Journal of Economic Perspectives*, 24(2), 31–46.

McFadden, D. (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlqvist, L. Lundsqvist, F. Snickars, and J. Weibull. North-Holland.

Miller, N. H., M. Remer, C. Ryan, and G. Sheu (2012): "Approximating the Price Effects of Mergers:Numerical Evidence and an Empirical Application," Working Paper.

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

Nevo, A., and M. Whinston (2010): "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference," *The Journal of Economic Perspectives*, 24(2), 69–82.

Reynolds, G., and C. Walters (2008): "The use of customer surveys for market definition and the competitive assessment of horizontal mergers," *Journal of Competition Law and Economics*, 4(2), 411–431.

Schmalensee, R. (2009): "Should New Merger Guidelines Give UPP Market Definition?," *Antitrust Chronicle*, 12, 1.

Sims, C. (2010): "But Economics Is Not an Experimental Science," *The Journal of Economic Perspectives*, 24(2), 59–68.

Stock, J. (2010): "The Other Transformation in Econometric Practice: Robust Tools for Inference," *The Journal of Economic Perspectives*, 24(2), 83–94.

Willig, R. (2011): "Unilateral Competitive Effects of Mergers: Upward Pricing Pressure, Product Quality, and Other Extensions," *Review of Industrial Organization*, 39(1-2), 19–38.

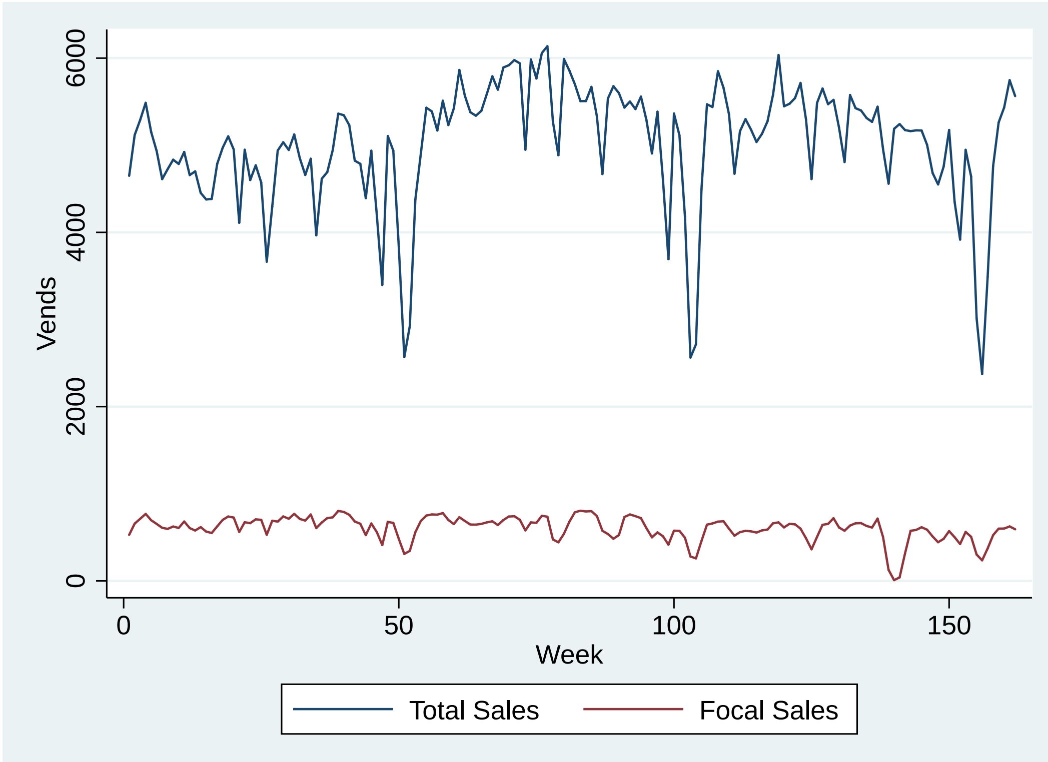

Figure 1: Total Overall Sales and Sales of Snickers and M&M Peanuts by Week

Table 1: Manufacturer Market Shares 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

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

Focal Sales	β_0	$\beta_{totalsales}$	R^2
Zoo Animal Cracker	6.424	0.039	0.503
Famous Amos Cookie	6.937	0.025	0.395
Doritos and Cheetos	-16.859	0.078	0.628
Snickers and Peanut M&Ms	-1.381	0.127	0.704

Table 2: Selection of Control Weeks: Regression for Focal Sales at Site 93

Experiment	Unadjusted	Adjusted	Adjusted (> 0)	γ
Zoo Animal Cracker	77.35	84.58	246.11	0.31
Famous Amos Cookie	65.30	-178.83	181.98	0.36
Cheetos and Doritos	66.22	-7.57	101.62	0.65
Snickers and Peanut M&M	66.47	17.69	79.00	0.84

Table 3: Adjustment for Aggregate Diversion

Manufacturer	Product	$Vends_0$	$Vends_1$	(% $Avail_0$)	(% $Avail_1$)	Adj Control	Difference
Hershey	Payday	1.2	6.6	1.3	2.5	2.3	4.3
Hershey	Reeses Peanut Butter Cups	56.5	174.7	29	72.5	141.3	33.4
Hershey	Twizzlers	24.3	57.3	17.7	37	50.7	6.6
Hershey	Choc Herhsey (Con)	37.2	65.9	29.9	35.5	44.2	21.7
Kellogg	Brown Sug Pop-Tarts	5.9	5.9	4.6	3.7	4.7	1.2
Kellogg	Strwbry Pop-Tarts	127.6	139.6	77.7	78.7	129.2	10.4
Kellogg	Cheez-It Original SS	190	206.2	84.6	82.3	184.8	21.4
Kellogg	Choc Chip Famous Amos	189.7	209.5	99.9	100	189.9	19.6
Kellogg	Rice Krispies Treats	26.1	82.7	30.7	60.8	51.7	31
Kraft	Ritz Bits Chs Vend	21.9	27.1	37	45.3	26.8	0.3
Kraft	100 Cal Oreo Thin Crisps	32.4	46.9	25.3	33.5	42.8	4.1
Mars	M&M Milk Chocolate	113.2	144.6	61	57.6	106.7	37.9
Mars	Milky Way	17	64.3	17.9	26.9	25.5	38.8
Mars	Twix Caramel	182.7	301.4	78.6	80	186	115.4
Mars	Nonchoc Mars (Con)	33.3	44.5	26.1	25.2	32.2	12.3
Nestle	Butterfinger	41.3	61.8	31.3	31.8	41.9	19.9
Nestle	Raisinets	149.2	199.3	80.8	83.1	153.5	45.8
Pepsi	Frito LSS	160.7	187.1	75.6	78.9	167.8	19.3
Pepsi	Grandmas Choc Chip	78.9	83.7	55.3	52	74.2	9.5
Pepsi	Lays Potato Chips 1oz SS	117.8	181.2	53.5	78.8	173.5	7.7
Pepsi	Baked Chips (Con)	208.1	228	91.7	97.5	221.3	6.7
Snyders	Snyders (Con)	367.6	418.6	82.6	87.2	387.9	30.7
Sherwood	Ruger Wafer (Con)	116.6	151.9	63.7	80.9	148.1	3.8
Kar's Nuts	Kar Sweet&Salty Mix 2oz	111.2	134.9	61.5	66.9	121.1	13.8
Misc	Rasbry Knotts	63.7	72.4	82.5	77.9	60.1	12.3
Misc	Farleys Mixed Fruit Snacks	66.6	78.8	57.3	67.4	78.3	0.5
Misc	Other Pastry (Con)	0.5	10.5	1.6	5	1.5	9
Misc	Salty Other (Con)	35.7	29.1	20.4	15.4	27.1	2

Table 4: Adjustments of Control Weeks: Mars (Snickers, Peanut M&M's) Experiment

Table 5: Parametric Model Estimates

	Random Coefficients		Nested Logit	
	Visit FE	Choice FE	Visit FE	Choice FE
σ_{Salt}	0.506	0.458		
	[.006]	[.010]		
σ_{Sugar}	0.673	0.645		
	[.005]	[.012]		
σ_{Peanut}	1.263	1.640		
	[.037]	[.028]		
$\lambda_{Chocolate}$			0.828	0.810
			[.003]	[.005]
$\lambda_{CandyNon-Choc}$			0.908	0.909
			[.007]	[.009]
$\lambda_{Cookie/Pastry}$			0.845	0.866
			[.004]	[.006]
λ_{Other}			0.883	0.894
			[.005]	[.006]
$\lambda_{SaltySnack}$			0.720	0.696
			[.003]	[.004]
# Nonlinear Params	3	3	5	5
Product FE	73	73	73	73
# Fixed Effects ξ_t	15,256	2,710	15,256	2,710
Total Parameters	15,332	2,786	15,334	2,788
LL	**-4,372,750**	**-4,411,184**	**-4,372,147**	**-4,410,649**
Total Sales	2,960,315	2,960,315	2,960,315	2,960,315
BIC	8,973,960	8,863,881	8,972,783	8,862,840
AIC	8,776,165	8,827,939	8,774,962	8,826,873

Manufacturer	Product	Adj. Diversion	Estimate(RC)	Estimate(NL)
Zoo Animal Cracker Experiment				
Mars	M&M Peanut	11.9	1.9	1.5
Mars	M&M Milk Chocolate	4.2	1.0	0.8
Mars	Snickers	7.6	1.8	1.4
Pepsi	Sun Chip LSS	4.2	0.9	0.9
Pepsi	Rold Gold (Con)	10.0	2.2	1.4
Famous Amos Experiment				
Hershey	Choc Herhsey (Con)	6.0	8.5	0.8
Pepsi	Sun Chip LSS	20.7	0.7	1.2
Pepsi	Rold Gold (Con)	9.1	1.3	1.8
Planters	Planters (Con)	10.5	0.6	1.8
Misc	Rasbry Knotts	3.9	0.3	1.3
Doritos and Cheetos Experiment				
Pepsi	Frito LSS	15.4	1.2	5.0
Pepsi	Baked Chips (Con)	6.3	1.1	4.8
Pepsi	FritoLay (Con)	11.1	1.4	3.7
Pepsi	Ruffles (Con)	7.3	1.3	6.0
General Mills	Nature Valley Swt&Salty Alm	3.9	0.6	0.5
Snickers and Peanut M&Ms Experiment				
Hershey	Reeses Peanut Butter Cups	4.1	4.1	3.1
Mars	M&M Milk Chocolate	4.7	4.0	3.6
Mars	Milky Way	4.8	2.4	2.5
Mars	Twix Caramel	14.2	6.4	4.3
Nestle	Raisinets	5.6	3.5	2.9

Table 6: Top 5 Substitutes by Experiment

Manufacturer	Differences	Adj. Differences	Adj. OLS	Nested Logit	RC Logit
Misc	23.6	8.3	11.3	6.7	5.8
Kar's Nuts	7.0	2.4	1.7	2.2	2.3
Sherwood	28.7	10.0	4.5	6.2	1.3
Pepsi	82.6	28.9	23.3	14.4	20.5
Nestle	0.2	0.1	4.4	2.0	3.4
Mars	6.1	2.2	12.7	9.6	15.6
Kraft	35.6	12.5	4.2	10.7	7.9
Kellogg	1.0	0.4	2.6	10.4	3.8
Hershey			.	2.6	4.2
Total	184.8	64.7	64.7	64.7	64.7
Misc	23.3	6.3	20.0	17.5	10.8
General Mills			.	2.0	1.3
Sherwood	19.2	5.2	.		
Pepsi	150.0	40.3	35.3	17.8	13.5
Mars			.	1.9	7.4
Kraft	34.8	9.4	5.6	8.7	8.7
Kellogg			.	12.4	7.4
Hershey	14.7	4.0	4.1	4.7	16.0
Total	241.9	65.1	65.1	65.1	65.1
Misc	8.2	4.1	13.0	10.4	11.7
General Mills	0.3	0.2	8.6	0.9	2.0
Sherwood	9.3	4.7	.		
Snyders	7.0	3.6	2.8		
Pepsi	94.4	47.7	37.8*	52.2	25.8
Mars			.	4.8	12.1
Kraft	8.2	4.1	1.9	2.2	6.0
Kellogg			.	1.1	3.1
Hershey	23.8	12.0	12.2	4.8	15.7
Total	151.2	76.4	76.4	76.4	76.4
Misc	3.6	3.1	1.7	3.2	1.6
Kar's Nuts	4.0	3.4	2.6	1.3	1.3
Sherwood	5.0	4.3	.	1.2	0.9
Snyders	0.6	0.6	2.3	3.5	2.1
Pepsi	1.4	1.2	8.9	6.6	6.0
Nestle	8.5	7.3	7.7**	8.7	8.5
Mars	25.2	21.7	9.7	19.1	18.7
Kraft	3.2	2.8	6.8*	2.5	4.7
Kellogg	1.9	1.7	10.5*	5.7	7.6
Hershey	20.6	17.8	13.8***	12.0	12.5
Total	74.1	63.9	63.9	63.9	63.9

Table 7: Diversion: Manufacturer Level Results

Manufacturer	Adj. Differences	Adj. OLS	Nested Logit	RC Logit
Animal Crackers Experiment				
Pepsi	18.6	18.3	9.2	13.1
Mars	2.3	16.6	10.3	16.7
Sherwood	6.3	2.2	3.9	0.8
Kraft	5.9	2.3	5.1	3.7
Kellogg	0.2	1.6	5.2	1.9
Hershey			3.1	5.0
Nestle	0.1	4.0	1.7	2.9
Famous Amos Cookie Experiment				
Pepsi	23.1	28.4	10.2	7.7
Hershey	4.2	3.1	5.0	16.9
Mars			1.8	7.1
Kellogg			5.5	3.3
Kraft	3.9	3.2	3.7	3.7
Cheetos and Doritos Nacho Experiment				
Pepsi	23.8	24.9	26.1	12.9
Hershey	11.1	7.7	4.4	14.5
Mars			4.0	10.1
General Mills	0.1	3.2	0.3	0.7
Snickers and M&M Peanut Experiment				
Mars	13.6	7.0	12.0	11.7
Hershey	12.3	5.8	8.3	8.6
Nestle	3.7	3.8	4.4	4.2
Pepsi	0.4	3.9	2.5	2.3
Kellogg	0.5	3.5	1.7	2.2

Table 8: Gross Upward Pricing Pressure Estimates

A Appendix:

A.1 diversion ratio under Parametric Demands

This section derives explicit formulas for the diversion ratio under common parametric forms for demand. The focus is whether or not a demand model implies that the diversion ratio is constant with respect to the magnitude of the price increase. It turns out that the IIA Logit and the Linear demand model exhibit this property, while the log-linear model, and mixed logit model do not necessarily exhibit this property.

We go through several derivations below:

Linear Demand
The diversion ratio under linear demand has the property that it does not depend on the magnitude of the price increase. To see this consider that the linear demand is given by:

$$Q_k = \alpha_k + \sum_j \beta_{kj} p_j$$

Which implies a diversion ratio corresponding to a change in price p_j of Δp_j:

$$D_{jk} = \frac{\Delta Q_k}{\Delta Q_j} = \frac{\beta_{kj}\Delta p_j}{\beta_{jj}\Delta p_j} = \frac{\beta_{kj}}{\beta_{jj}} \tag{18}$$

This means that for any change in p_j from an infinitesimal price increase, up to the choke price of j; the diversion ratio, D_{jk} is constant. This also implies that under linear demands, divergence is a global property, under any initial set of prices, quantities, or any magnitude of price increase will result in the same diversion.

Log-Linear Demand
The log-linear demand model does not exhibit constant diversion with respect to the magnitude of the price increase. The log-linear model is specified as:

$$\ln(Q_k) = \alpha_k + \sum_j \epsilon_{kj} \ln(p_j)$$

If we consider a small price increase Δp_j the diversion ratio becomes:

$$\begin{aligned}\frac{\Delta \log(Q_k)}{\Delta \log(Q_j)} \approx \underbrace{\frac{\Delta Q_k}{\Delta Q_j}}_{D_{jk}} \cdot \frac{Q_j(\mathbf{p})}{Q_k(\mathbf{p})} &= \frac{\varepsilon_{kj}\Delta \log(p_j)}{\varepsilon_{jj}\Delta \log(p_j)} = \frac{\varepsilon_{kj}}{\varepsilon_{jj}} \\ D_{jk} &\approx \frac{Q_k(\mathbf{p})}{Q_j(\mathbf{p})} \cdot \frac{\varepsilon_{kj}}{\varepsilon_{jj}} \end{aligned} \tag{19}$$

This holds for small changes in p_j. However for larger changes in p_j we can no longer use the simplification that $\Delta \log(Q_j) \approx \frac{\Delta Q_j}{Q_j}$. So for a large price increase (such as to the choke price $p_j \to \infty$, log-linear demand can exhibit diversion that depends on the magnitude of the price increase.

IIA Logit Demand
The plain logit model exhibits IIA and proportional substitution. This implies that the diversion ratio does not depend on the magnitude of the price increase. Here we consider two price increases, an infinitesimal one and an increase to the choke price $p_j \to \infty$.

Consider the derivation of the diversion ratio D_{jk} under simple IIA logit demands. We have utilities and choice probabilities given by the well known equations, where a_t denotes the set of products available in market t:

$$\begin{aligned} u_{ijt} &= \underbrace{x_{jt}\beta - \alpha p_{jt}}_{\tilde{v}_{jt}} + \varepsilon_{ijt} \\ S_{jt} &= \frac{\exp[\tilde{v}_{jt}]}{1 + \sum_{k \in a_t} \exp[\tilde{v}_{kt}]} = \frac{V_{jt}}{IV(a_t)} \end{aligned}$$

Under logit demands, an infinitesimal price change in p_1 exhibits identical diversion to setting $p_1 \to \infty$ (the choke price):

$$\begin{aligned} \widehat{D_{jk}} &= \frac{\frac{\partial S_k}{\partial p_j}}{\left|\frac{\partial S_j}{\partial p_j}\right|} = \frac{\alpha S_k S_j}{\alpha S_j (1 - S_j)} = \frac{S_k}{(1 - S_j)} \\ \overline{D_{jk}} &= \frac{\frac{e^{V_k}}{1 + \sum_{l \in a \setminus j} e^{V_l}} - \frac{e^{V_k}}{1 + \sum_{l' \in a} e^{V_l'}}}{0 - \frac{e^{V_j}}{1 + \sum_{l \in a} e^{V_l}}} = \frac{S_k}{(1 - S_j)} \end{aligned}$$

As an aside $\frac{S_k}{1-S_k} = \frac{Q_k}{M-Q_k}$, so we either need market shares or market size (back to market definition!). In both cases diversion is merely the ratio of the marketshare of the substitute good divided by the share not buying the focal good (under the initial set of prices and product availability). It does not depend on any of the econometric parameters (α, β).

Also we can also show that the bias expression for the diversion ratio is set to zero that is: $D_{jk} = \frac{\partial^2 q_k}{\partial p_j^2} / \frac{\partial^2 q_j}{\partial p_j^2}$.

$$\begin{aligned} \frac{\partial^2 q_j}{\partial p_j^2} &= \alpha^2 (1 - 2S_j)(S_j - S_j^2) \\ \frac{\partial^2 q_k}{\partial p_j^2} &= -\alpha^2 (1 - 2S_j) S_j S_k \\ \frac{\frac{\partial^2 q_k}{\partial p_j^2}}{\frac{\partial^2 q_j}{\partial p_j^2}} &= \frac{S_k}{1 - S_j} = D_{jk} \end{aligned}$$

Random Coefficients Logit Demand

Random Coefficients Logit demand relaxes the IIA property of the plain Logit model, which can be undesirable empirically, but it also means that the diversion ratio varies with original prices and quantities, as well as with the magnitude of the price increase. Intuitively a small price increase might see diversion from the most price sensitive consumers, while a larger price increase might see substitution from a larger set of consumers. If price sensitivity is correlated with other tastes, then the diversion ratio could differ with the magnitude of the price increase.

We can repeat the same exercise for the logit model with random coefficients, by discretizing a mixture density over $i = 1, \ldots, I$ representative consumers, with population weight w_i:

$$u_{ijt} = \underbrace{x_{jt}\beta_i - \alpha_i p_{jt}}_{\tilde{v}_{ijt}} + \varepsilon_{ijt}$$

Even when consumers have a common price parameter $\frac{\partial V_{ik}}{\partial p_j} = \alpha$,

$$\widehat{D_{jk}} = \frac{\frac{\partial S_k}{\partial p_j}}{\left|\frac{\partial S_j}{\partial p_j}\right|} = \frac{\int s_{ij} s_{ik} \frac{\partial V_{ik}}{\partial p_j}}{\int s_{ij}(1-s_{ij})\frac{\partial V_{ij}}{\partial p_j}} \to \frac{\int s_{ij} s_{ik}}{\int s_{ij}(1-s_{ij})}$$

$$\overline{D_{jk}} = \frac{\int \frac{e^{V_{ik}}}{1+\sum_{l\in a\setminus j} e^{V_{il}}} - \frac{e^{V_{ik}}}{1+\sum_{l'\in a} e^{V_{il'}}}}{\int -\frac{e^{V_{ij}}}{1+\sum_{l\in a} e^{V_{il}}}} = \frac{1}{s_j}\int \frac{s_{ij} s_{ik}}{(1-s_{ij})}$$

Now, each individual exhibits constant diversion, but weights on individuals vary with p, so that diversion is only constant if $s_{ij} = s_j$. Otherwise observations with larger s_{ij} are given more weight in correlation of $s_{ij}s_{ik}$. The more correlated (s_{ij}, s_{ik}) are (and especially as they are correlated with α_i) the greater the discrepancy between marginal and average diversion.