An Experimental Approach to Merger Evaluation *

Christopher T. Conlon†
Julie Holland Mortimer‡

(Click Here for Current Version)

December 30, 2016

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), which relies on a *diversion ratio* that measures the fraction of consumers of one product that switch to another product when the price of the first product increases. Typically the diversion ratio is computed by estimating own- and cross-price elasticities from a demand system. We show that it is possible to reinterpret the diversion ratio as the treatment effect of removing a product from the consumer's choice set. We derive conditions on economic primitives under which one can obtain accurate estimates of the treatment effect in the presence of unobserved demand shocks, and conditions under which this treatment effect accurately represents the diversion ratio. We demonstrate our approach in a field experiment on snack foods.

*We thank Mark Stein, Bill Hannon, and the drivers at Mark Vend Company for implementing the experiments used in this paper, providing data, and generally educating us about the vending industry. 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.

†NYU Stern, Henry Kaufman Management Center, 44 West Fourth Street, New York, NY 10012. email: cconlon@stern.nyu.edu

‡Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467, and NBER. email: 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 an index of market concentration known as the Herfindahl-Hirschman Index (HHI). The HHI relates market shares to markups when firms are engaged in Cournot competition. A practical challenge of using the HHI to evaluate merger effects has been defining the relevant market, which is necessary for measuring market shares. The proposed 2007 merger between *Whole Foods* and *Wild Oats* highlighted this challenge: the FTC argued that the merger would create a monopoly in the market for "premium natural organic supermarkets," while Whole Foods argued that the relevant market included traditional grocery stores, and that the merger would induce only a minimal change in market concentration.[1]

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. The new guidelines ignore competitive responses and consider the 'unilateral effects' of a proposed merger on the opportunity cost of increasing the price of the merged entity's products. The approach relies on two key measures: Upward Pricing Pressure (UPP) and a Generalized Upward Pricing Pressure Index (GUPPI), both of which depend on the prices and costs of the merged firms' products, and a 'diversion ratio,' which measures substitution between the products of the merged firms. The diversion ratio thus 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 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*

[1] The merger initially was permitted on the basis of the latter market definition, but then dissolved on the basis of the former.

effects. Diversion ratios between products sold by merging firms and those sold by non-merging firms have at most secondary predictive value.

Thus, the unilateral effects approach outlined by the 2010 Horizontal Merger Guidelines implies that the diversion ratio functions as a sufficient statistic (holding competitive responses fixed); and we should be more suspicious of mergers among products with higher diversion ratios. Accordingly, we focus on measuring substitution away from product j (from second-choice data or stock-outs) and ask, 'Is there a credible (quasi)-experimental way to measure the diversion ratio, and if so, what assumptions does it require?' To answer this question, we show that the diversion ratio can be interpreted as a treatment effect of an experiment in which the treatment is "not purchasing product j." The diversion ratio measures the outcome of the experiment: the fraction of consumers who switch from j to a substitute product k. The treated group consists of consumers who would have purchased j at pre-existing prices, but no longer purchase j at a higher price. We are often interested in measuring the effect when the treated group consists of only those consumers who are most likely to substitute away from j after a very small price increase. This effect represents a marginal treatment effect (MTE) or a local average treatment effect (LATE). A challenge with directly implementing such an experiment is that treating a small number of the most price-sensitive individuals lacks statistical power.

An alternative is to treat all individuals who would have purchased j at pre-existing prices, and thus estimate an average treatment effect (ATE) when all individuals receive the treatment. This can be accomplished by surveying consumers about their second-choice products, or by exogenously removing product j from the choice set. When the diversion ratio is constant, the ATE approximates the MTE. However, we show that constant diversion is a feature of only the linear demand model and a "plain vanilla" logit model. Other commonly-used models of demand, such as a random-coefficients logit or log-linear models , do not feature constant diversion, and the ATE and MTE may diverge in important ways in these models.

We construct an empirical estimator for the ATE measure of the diversion ratio by exogenously removing individual products from a local market; specifically, a set of vending machines. Even when product removals are appropriately randomized, it can be difficult to control for variability in the overall level of demand. This is true in our setting, and is likely to be true in many retail contexts in which we wish to measure diversion. Thus, one must separate the effect of the exogenous removal from unobserved demand shocks. In order to control for unobserved demand shocks, we provide three conditions on economic primitives and examine how they help to estimate experimental measures of the diversion ratio. The conditions are: (1) product removals cannot increase total sales, nor decrease total sales by more than the sales of the product removed, (2) the diversion ratio from j to any substitute k lives within the unit interval (i.e., that diversion to any single product is between 0 and 100 percent), and (3) the diversion ratio to all alternatives lives within the unit simplex (i.e., that total diversion to all alternatives is between 0 and 100 percent). We construct a non-parametric Empirical Bayes shrinkage estimator for the diversion ratio, which allows us to nest both parametric structural estimates of diversion and (quasi)-experimental measures in a single framework.

Our empirical example demonstrates how to design and conduct experiments to measure the diversion ratio, and considers several hypothetical mergers within the single-serving snack foods industry. 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. 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.

1.1 Related Literature

There has been a recent debate on the use of experimental or quasi-experimental techniques vis-a-vis structural methods within industrial organization (IO) broadly, and within merger evaluation specifically. Angrist and Pischke (2010) complain about the general lack of experimental or quasi-experimental variation in many IO papers, and advocate viewing a merger itself as the treatment effect of interest. Hastings (2004) is cited as an example, which examines the effect of a merger between Thrifty and ARCO gasoline stations on the prices of competitors near and far away from affected stations. Nevo and Whinston (2010)'s response points out that while retrospective merger analysis is valuable, the policy question is generally one of prospective merger analysis, and that merely comparing proposed mergers to similar previously consummated mergers is unlikely to be informative, especially when both the proposal and approval of mergers is endogenous. They point out that while labor economics is often concerned with a single treatment effect, many of the key issues in IO are concerned with testing the equilibrium implications of economic theory, and that the complex competitive responses that arise in market settings often do not map well into the treatment effects framework.[2]

A deeper question, posed by the applied theory literature in IO, focuses on whether or not the diversion ratio is likely to be informative about the price effect of a merger in the first place. This literature goes back as far as Shapiro (1995) and Werden (1996), and is well summarized in reviews by Farrell and Shapiro (2010) and Werden and Froeb (2006). A growing debate has developed out of the expanded role of unilateral effects in the 2010 Horizontal Merger Guidelines including: Carlton (2010), Schmalensee (2009), Willig (2011), and Hausman (2010). The UPP approach measures the change in the opportunity cost of selling good j that is induced by a merger.[3] There have also been a few recent attempts to validate the predictions of the unilateral effects approach in simulation (Miller, Remer, Ryan, and Sheu 2012) and empirically (Cheung 2011). Those papers find that the UPP/GUPPI need not always correctly predict the sign of the price effect, and that the degree to which it over- or underpredicts price effects depends on the nature of competition among non-merging firms, and whether prices are strategic substitutes or strategic complements.

In spirit, our approach is most similar to Angrist, Graddy, and Imbens (2000), which shows how a cost shock can identify a particular local average treatment effect (LATE) for the price elasticity in a single product setting. However, that approach does not extend to the differentiated products setting because the requisite monotonicity condition may no longer be satisfied. While the diversion ratio is often constructed from the ratio of own- and cross price elasticities, our approach estimates the (diversion) ratio directly. Though own- and cross- price elasticities are not identified from second-choice data alone, the average diversion ratio is. This highlights the economic content of (even partial) second-choice data, which has been found to be valuable in the structural literature on demand estimation (Berry, Levinsohn, and Pakes 2004).

Finally, by asking the question 'What assumptions does a credible (quasi-experimental)

[2]This is actually a reexamination of a much older debate going back to Leamer (1983), and discussed recently by Heckman (2010), Leamer (2010), Keane (2010), Sims (2010), Stock (2010), and Einav and Levin (2010).

[3]Jaffe and Weyl (2013) extend these results by incorporating the estimated pass-through rate to map anticipated opportunity cost effects into price effects.

method of measure diversion require,' we connect directly to the nascent theoretical literature discussing the use and measurement of the diversion ratio.[4] Farrell and Shapiro (2010) suggest that firms themselves track diversion in their 'normal course of business',' or that the diversion ratio is essentially another piece of data likely to be uncovered in a Hart-Scott-Rodino filing. Hausman (2010) argues that the only acceptable way to measure a diversion ratio is as the output from a structural demand system. Reynolds and Walters (2008) examine the use of stated-preference consumer surveys in the UK for measuring diversion.

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 and the calculation of the treatment effects. We present the results of our field experiment in section 4, and section 5 concludes.

2 Theoretical Framework

The first part of this section establishes results presented in Farrell and Shapiro (2010), using slightly different notation to define the key constructs of the 2010 merger guidelines. The alternative notation aids in our treatment-effects interpretation.

For simplicity, consider a single market composed of J single-product firms, where firm j sets the price of product j to maximize profits:

$$\pi_j = (p_j - c_j(q_j))q_j(p_j, p_{-j})$$

Under an assumption of constant marginal costs, the FOC for product j becomes

$$q_j(p_j, p_{-j}) + (p_j - c_j)\frac{\partial q_j(p_j, p_{-j})}{\partial p_j} = 0$$

Let the superscripts (0) and (1) denote pre- and post-merger quantities respectively. The merger guidelines also consider the potential for a merger to induce efficiency gains by lowering the cost of producing product j. This efficiency gain, denoted e_j, is defined as:

$$e_j = \frac{c_j^{(1)} - c_j^{(0)}}{c_j^{(0)}}.$$

[4]The focus on measuring substitution away from product j (using second-choice data or stock-outs), rather than on the direct effect of a proposed merger, is more in line with the public finance literature on sufficient statistics (Chetty 2009).

Henceforth, we denote pre-merger costs as c_j and post-merger costs as $(1-e_j)\cdot c_j$. We are interested in how a merger affects the price of j : $p_j^{(1)} - p_j^{(0)}$. The idea is to examine how the merger modifies the FOC of a single-product firm which owns j and acquires product k, when prices of all other goods p_{-j} are held fixed at the pre-merger values:

$$
\begin{aligned}
q_j(p_j^{(0)}, p_{-j}) + (p_j^{(0)} - c_j)\frac{\partial q_j(p_j^{(0)}, p_{-j})}{\partial p_j} & = 0 \quad \text{(Pre-merger)} \\
q_j(p_j^{(1)}, p_{-j}) + (p_j^{(1)} - (1-e_j)\cdot c_j)\frac{\partial q_j(p_j^{(1)}, p_{-j})}{\partial p_j} + \quad (p_k - c_k)\cdot \frac{\partial q_k(p_j^{(1)}, p_{-j})}{\partial p_j} & = 0 \quad \text{(Post-merger)}
\end{aligned}
$$

Upward Pricing Pressure (UPP) for good j measures how the merger changes the FOC for good j holding p_{-j} fixed. The intuition is that a merger changes the opportunity cost of selling j, because some lost sales are now recaptured by product k. We compare the change in the opportunity cost of selling product j to the proposed efficiency gain e_j in order to determine whether a merger is likely to increase or decrease the price of j.

$$
UPP_j = (p_k - c_k)\cdot \underbrace{\left(\frac{\partial q_j(p_j^{(0)}, p_{-j})}{\partial p_j}\right)^{-1} \cdot \frac{\partial q_k(p_j^{(0)}, p_{-j})}{\partial p_j}}_{D_{jk}(p_j^{(0)}, p_{-j})} - e_j \cdot c_j \tag{1}
$$

The key input into merger analysis is the diversion ratio, $D_{jk}(p_j, p_{-j})$. The diversion ratio measures the fraction of consumers who switch from j to k when faced with a small increase in the price of j from the pre-merger prices $(p_j^{(0)}, p_{-j})$. The 2010 merger guidelines also define a measure of pricing pressure called the Gross Upward Pricing Pressure Index (GUPPI), which (i) assumes no marginal cost efficiency e_j and (ii) indexes the change in price to the pre-merger price level. Thus, GUPPI takes the form:

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

If the owner of j acquires multiple products k and l, and those products have the same margins $p_k - c_k = p_l - c_l$, then the impact of the three-product merger on UPP_j and $GUPPI_j$ is identical to that in equations (1) and (2), with the exception that it depends on the sum of the diversion ratios $(D_{jk} + D_{jl})$. This allows one to incorporate the acquisition of multiple flavors or similar brands. It also provides a template for considering how various divestitures might impact a proposed merger.

2.1 Diversion as a Treatment Effect

The key to the diversion ratio is that it holds the prices (and competitive responses) of all other goods fixed at p^0_{-j} and considers what happens when one perturbs p_j by Δp_j:

$$D_{jk}(p_j, p^0_{-j}) = \left|\frac{\Delta q_k}{\Delta q_j}\right| = \left|\frac{q_k(p^0_j + \Delta p_j, p^0_{-j}) - q_k(p^0_j, p^0_{-j})}{q_j(p^0_j + \Delta p_j, p^0_{-j}) - q_j(p^0_j, p^0_{-j})}\right| = \frac{\int_{p^0_j}^{p^0_j+\Delta p_j} \frac{\partial q_k(p_j, p^0_{-j})}{\partial p_j}\, dp_j}{\int_{p^0_j}^{p^0_j+\Delta p_j} \frac{\partial q_j(p_j, p^0_{-j})}{\partial p_j}\, dp_j} \tag{3}$$

This has the interpretation as a treatment effect with a binary treatment (i.e., not purchasing product j) and a binary outcome (i.e., purchasing product k or not). The treated group corresponds to individuals who would have purchased product j at price p_j but do not purchase j at price $p_j + \Delta p_j$. The "potential outcome" $Y_i(1) \in \{0, 1\}$ indicates whether or not a treated individual purchases product k. The lower an individual's reservation price for j, the more likely an individual is to receive the treatment. Thus Δp_j functions as the "instrument" in this context.

By focusing on the numerator in equation (3), we can re-write the diversion ratio using the marginal treatment effects (MTE) framework of Heckman and Vytlacil (2005), in which $D_{jk}(p_j, p^0_{-j})$ is a marginal treatment effect that depends on p_j.[5]

$$\widehat{D_{jk}^{LATE}} = \frac{1}{\Delta q_j} \int_{p^0_j}^{p^0_j+\Delta p_j} \underbrace{\frac{\partial q_k(p_j, p^0_{-j})}{\partial q_j}}_{\equiv D_{jk}(p_j, p^0_{-j})} \left|\frac{\partial q_j(p_j, p^0_{-j})}{\partial p_j}\right| dp_j \tag{4}$$

As we vary p_j, we measure the weighted average of diversion ratios where the weights $w(p_j) = \frac{1}{\Delta q_j} \frac{\partial q_j(p_j, p^0_{-j})}{\partial p_j}$ correspond to the lost sales of j at a particular p_j as a fraction of all lost sales. This directly corresponds to Heckman and Vytlacil (2005)'s expression for the local average treatment effect (LATE); we average the diversion ratio over the set of consumers of product j who are most price sensitive. Our LATE estimator varies because the set of treated individuals varies with the size of the price increase. The expression for UPP in equation (1) evaluates $D_{jk}(p^0_j, p^0_{-j})$ at pre-merger prices. This is consistent with a MTE for which Δp

[5]The MTE is a non-parametric object which can be integrated over different weights to obtain all of the familiar treatment effects estimators: treatment on the treated, average treatment effects, local average treatment effects, average treatment on the control, etc.

is infinitessimally small.[6] As we choose larger values for Δp_j our LATE estimate may differ from the MTE.

We can relate the divergence in the treatment effect measures of D_{jk} to the underlying economic primitives of demand. Consider what happens when we examine a "larger than infinitesimal" increase in price $\Delta p_j \gg 0$. We derive an expression for the second-order expansion of demand at (p_j, p_{-j}):

$$\begin{aligned} q_k(p_j + \Delta p_j, p_{-j}) &\approx q_k(p_j, p_{-j}) + \frac{\partial q_k(p_j, p_{-j})}{\partial p_j}\Delta p_j + \frac{\partial^2 q_k(p_j, p_{-j})}{\partial p_j^2}(\Delta p_j)^2 + O((\Delta p_j)^3) \\ \frac{q_k(p_j + \Delta p_j, p_{-j}) - q_k(p_j, p_{-j})}{\Delta p_j} &\approx \frac{\partial q_k(p_j, p_{-j})}{\partial p_j} + \frac{\partial^2 q_k(p_j, p_{-j})}{\partial p_j^2}\Delta p_j + O(\Delta p_j)^2 \end{aligned} \tag{5}$$

This allows us to compute an expression for the bias compared to the true diversion ratio $D_{jk}(p_j, p_{-j})$

$$Bias(D_{jk}^{LATE}) \approx -\frac{D_{jk}\frac{\partial^2 q_j}{\partial p_j^2} + \frac{\partial^2 q_k}{\partial p_j^2}}{\frac{\partial q_j}{\partial p_j} + \frac{\partial^2 q_j}{\partial p_j^2}\Delta p_j}\Delta p_j \tag{6}$$

The expression in (6) shows that the bias depends on two things: one is the magnitude of the price increase Δp_j, the second is the curvature of demand (the terms $\frac{\partial^2 q_j}{\partial p_j^2}$ and $\frac{\partial^2 q_k}{\partial p_j^2}$). This suggests that bias is minimized by experimental designs that consider small price changes.

The disadvantage of considering a small price change Δp_j is that it implies that the size of the treated group Δq_j is also small, and thus the variance of our diversion measure is large. We can construct an expression for the variance of the diversion ratio under the assumption of (locally) constant diversion, for which $\Delta q_k \approx D_{jk}\Delta q_j$:

$$Var(\widehat{D_{jk}}) \approx Var\left(\frac{\Delta q_k}{|\Delta q_j|}\right) \approx \frac{1}{\Delta q_j^2}\left(D_{jk}^2\sigma_{\Delta q_j}^2 + \sigma_{\Delta q_k}^2 - 2D_{jk}\rho\sigma_{\Delta q_j}\sigma_{\Delta q_k}\right) \tag{7}$$

This expression establishes a bias-variance tradeoff when estimating diversion. A small change in p_j induces a small change in q_j and reduces the bias, but increases the potential variance. A larger Δp_j (and by construction Δq_j) may yield a less noisy LATE, but may deviate from the MTE of interest.

Because the underlying objects are essentially demand curves, economic theory provides

[6] Anti-trust authorities also sometime sfocus on the notion of a 'small but significant non-transitory increase in price (SSNIP).' The practice of antitrust often employs an SSNIP test of 5-10%.

some guidance. A key question is: What are the economic implications of assuming a constant treatment effect, such that $D_{jk}(p_j, p_{-j}) = D_{jk}$? We can see the answer by examining the case where (6) is equal to zero. Two functional forms for demand 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 $D_{jk} = -\frac{\partial^2 q_k}{\partial p_j^2} / \frac{\partial^2 q_j}{\partial p_j^2}$. *Implicitly when we assume that the diversion ratio does not vary with price, we assume that the true demand system is well approximated by either a linear demand curve or the IIA logit model.* We derive these relationships, as well as expressions for diversion under other demand models in the Appendix A.1, and show that random coefficients logit demand, and CES demands (including log-linear demand) do not generally exhibit constant diversion.

If our primary concern is that the curvature of demand is steep, so that assuming a constant diversion ratio is unreasonable, it suggests considering a small price increase. However, if our primary concern is that sales are highly variable, we may need to consider a larger price increase. Information about the elasticity (and super-elasticity) of demand for j can be very informative. A LATE estimate is expected to perform best if demand for j is highly elastic around (p_j, p_{-j}), or if demand becomes increasingly inelastic as we increase the price beyond p_j.

In our empirical 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. In Figure 2, 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 after the first few thousand dollars of price increases, then assuming constant diversion may be reasonable (because the implicit weight on faraway diversion measures is small). If demand for the Prius is relatively inelastic or does not become more inelastic as the price rises (the dashed green line), then assuming a constant treatment effect may lead to bias.[7]

[7]Mathematically, the average diversion ratio represents the integral of the diversion curve (red line) from market price upwards, where the diversion curve is weighted by the density given by the demand for the Prius (either green line).

2.2 Second-Choice Data

Often researchers have access to one form or another of second-choice data. For example, Berry, Levinsohn, and Pakes (2004) observed not only marketshares of cars but also survey answers to the question: "If you did not purchase this vehicle, which vehicle would you purchase?" Consumer surveys provide a stated-preference method of recovering second-choice data. The UK Competition Commission makes use of consumer surveys as part of the merger review process both for market definition, and for second-choice data ((Reynolds and Walters 2008)).

Exploiting variation in consumer choice sets provides a revealed-preference mechanism for recovering second-choice data. A problem with using observational variation in choice sets is that the variation is often non-random. If one simply compares stores that stock product j to stores that do not stock product j, one might expect the stocking decision to be correlated with demand for both j and other products. In previous work, Conlon and Mortimer (2013) establish conditions under which a temporary stock-out event provides random variation in the choice set. The main intuition is that given inventory decisions and consumer demand, the timing of a stock-out is randomly distributed, paired with the assumption that consumer arrival patterns do not respond to anticipated stock-out events; this provides random choice set variation.

A more direct approach is to construct second-choice data experimentally by removing product j from a consumer's choice set for a period of time. One way to interpret second-choice data or an experimental product removal is as an increase in price to the choke price p_j^c, where $q_j(p_j^c, p_{-j}) = 0$. The precise location of the choke price varies with the nature of demand. For linear demand $q_j = a + bp_j$ the choke price is $p = \frac{-a}{b}$. For random utility demand models with full-support errors, the choke price is $p_j = \infty$. In practice, the quantity estimated by a random product removal (or by second-choice data generally) is the average treatment effect on the treated (ATT) for the entire population with reservation prices between $p_j^{(0)}$ and p_j^c.[8] One advantage of the product removal experiment is that it treats as many individuals as possible, and thus minimizes the variance expression in (7).

The ATE provides a good approximation for the MTE when the expression in (6) is small: (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) $D_{jk}(p_j, p_{-j}) = D_{jk}$, or (c) when demand for j is

[8]One could take this to correspond to the entire population of individuals who might substitute away from j when faced with a price increase. It does not include those individuals who might substitute towards j given a price *decrease*. Given valid randomization, we would expect that $ATE \approx ATT$ because the entire demand curve is treated.

steepest near the market price $\left|\frac{\partial q_j(p_j, p_{-j})}{\partial p_j}\right| \gg \left|\frac{\partial q_j(p_j+\Delta p_j, p_{-j})}{\partial p_j}\right|$. A limitation is that unless we know the true functional form of demand, we cannot derive an explicit form for the discrepancy between the ATE we measure and the MTE of interest. In Appendix A.2 we conduct some simulations with commonly used parametric demand models, and report the maximum discrepancy between the MTE and the ATE estimates.

2.3 Empirical Approach

In many retail settings, variability in the overall level of demand may overwhelm any experimentally-induced variation from a small price change. Thus, our empirical approach focuses on estimating the ATE using experimentally-generated second-choice data. Our objective is to obtain an accurate estimate of the ATE under a minimal set of parametric assumptions, while still controlling for overall variability in the level of demand.

There are a discrete number of treated individuals Δq_j who would have purchased j had it been available in the control but cannot in the treatment, and a discrete number of individuals Δq_k, who purchase k in the treatment but would not during the control. Because both outcome and treatment are the sum of discrete events, we can express the diversion ratio using the binomial distribution with $\Delta q_k \sim Bin(n = \Delta q_j, p = D_{jk})$. Estimating the diversion ratio merely requires estimating the binomial parameter $\hat{p} = \frac{\Delta q_k}{\Delta q_j}$ which is straightforward.

The challenge arises because both $\Delta q_j, \Delta q_k$ require assumptions on potential outcomes.

$$\begin{aligned}\Delta q_j = &\ \underbrace{E[q_j|T=1, X=x, k \text{ available}]}_{=0} - &\ E[q_j|T=0, X=x, k \text{ available}] \\ \Delta q_k = &\ E[q_k|T=1, X=x, k \text{ available}] - &\ E[q_k|T=0, X=x, k \text{ available}]\end{aligned}$$

One advantage of using product removal experiments is that $E[q_j|T=1, X=x] = 0$ by construction (consumers cannot purchase products that are unavailable). This also helps rule out one set of potential defiers. The second set of defiers, those that purchase k only when j is available are ruled out if j, k are substitutes rather than complements.

Even when we have random assignment of the treatment T, we still need to ensure that the distribution of covariates $f(x)$ is the same for treatment and control periods. Because we are interested in the ratio of treatment effects, we also need to ensure that $f(x)$ is the same for both Δq_j and Δq_k. This suggests a *matching* or *balancing* approach given random assignment of T. The most obvious covariate is that product k (the substitute) is actually available to consumers. The second covariate we should worry about is the overall level of demand. The overall level of demand may be higher (or lower) during the treatment period than the corresponding period from the control. This presents two additional challenges: the first is that there may not be a single scalar x for overall demand that affects all products in the same way, instead we might have product specific shocks (x_j, x_k); the second is that we don't expect to directly observe x even if it is scalar.

The approach we propose is to derive weaker conditions that help to balance the treatment and control periods without observing x directly. Our first assumption is a weak implication of all products being substitutes for one another.[9]

Assumption 1. *"Substitutes": Removing product j can never increase the overall level of sales during a period, and cannot decrease sales by more than the sales of j.*

We let q_{jt} denote the sales of product j during period t, and Q_t denote the sales of all products during period t (not just products (j,k)). Given a treatment period t, we look for the corresponding set of control periods which satisfy Assumption 1:

$$\{s : s \neq t, T = 0, k \text{ avail }, Q_s - Q_t \in [0, q_{js}]\} \tag{8}$$

The problem with a direct implementation of (8) is that periods with higher sales of the focal product q^0_{js} are more likely to be included in the control, which would understate the diversion ratio. We propose a slight modification of (8) which is unbiased. We replace q_{js} with $\widehat{q_{js}} = E[q_{js}|Q_s, T = 0]$. An easy way to obtain the expectation is to run an OLS regression of q_{js} on Q_s using data only from control periods for which j is available.

$$S_t = \{s : s \neq t, Q^0_s - Q^1_t \in [0, \widehat{b_0} + \widehat{b_1} Q^0_s]\} \tag{9}$$

Thus (9) defines the set of control periods S_t which correspond to treatment period t under our assumption. The economic implication of Assumption 1 is that the sum of the diversion ratios from j to all other products is between zero and one (for each t): $\sum_{k \neq j} D_{jk,t} \in [0, 100\%]$.

We might be willing to make the additional assumption that each individual diversion ratio is between zero and one: $\forall k, D_{jk} \in [0, 100\%]$. In fact, the interpretation of the diversion

[9] There are stronger assumptions we could make in order to implement a more traditional matching or balancing estimator in the spirit of Abadie and Imbens (2006). Suppose a third product k' was similarly affected by the demand shock x but we knew ex-ante that $D_{jk'} = 0$, we could match on similar sales levels of $q_{k'}$. For our vending example this might be using sales at a nearby soft drink machine to control for overall demand at the snack machine, or it might be using sales of chips to control for sales of candy bars.

ratio as the binomial parameter implicitly requires $D_{jk} \in [0,1]$. We may also be interested in comparing diversion ratios across substitutes k, k'. One challenge may be that for some k and t the corresponding set S_t may be small (or even empty), while for some k' the corresponding set may be large. In one case, we have very little information about D_{jk} and the other we have fairly precise information about $D_{jk'}$. We might erroneously conclude that D_{jk} is large and suggest the merging parties divest k when in fact it is not an issue. One way to address both the precision of the diversion estimates, and constrain them to $\in [0,1]$, is to impose a prior distribution on D_{jk}.

Assumption 2. *"Unit Interval": $D_{jk} \in [0,1]$. $\Delta q_k | \Delta q_j, D_{jk} \sim Bin(n = \Delta q_j, p = D_{jk})$ and $D_{jk}|\beta_1, \beta_2 \sim Beta(\beta_1, \beta_2)$.*

This assumption implies that the posterior distribution is given by $D_{jk}|\beta_1, \beta_2, \Delta q_j, \Delta q_k \sim Beta(\beta_1 + \Delta q_k, \beta_2 + \Delta q_j - \Delta q_k)$, and has posterior mean $E[D_{jk}|\beta_1, \beta_2, \Delta q_j, \Delta q_k] = \frac{\beta_1 + \Delta q_k}{\beta_1 + \beta_2 + \Delta q_j}$. An alternate interpretation of the beta-binomial conjugacy is that we observe Δq_j treated individuals and Δq_k "successes" but wish to incorporate prior information about the probability of the success from similar trials observed outside our dataset. The mean of the prior is given by $\mu_{jk} = \frac{\beta_1}{\beta_1 + \beta_2}$, and the weight we put on the prior is equivalent to having $m_{jk} = \beta_1 + \beta_2$ "pseudo-observations" in the data, that were observed before any experimental data. We can also write our estimate for the mean of D_{jk} as a shrinkage estimator:

$$D_{jk} = \lambda \cdot \mu_{jk} + (1 - \lambda)\frac{\Delta q_k}{\Delta q_j}, \quad \lambda = \frac{m_{jk}}{m_{jk} + \Delta q_j}$$

Thus λ tells us how much weight to put on our prior mean versus our experimental observations, or how many "pseudo-observations" we observed of our prior before we saw our experimental outcomes. One reason this estimator is referred to as a "shrinkage" estimator, is because as Δq_j becomes smaller (and our experimental outcomes are less informative), $\hat{D}_{jk}$ is shrunken towards μ_{jk}. Thus, when our experiments provide lots of information about diversion from j to k we use rely on the experimental outcomes, but when our experiments are less informative we rely more on our prior information.

The challenge in any Bayesian analysis is how to choose the prior mean μ_{jk}. Here we have strong guidance both from economic theory and the practice of antitrust. One possible choice for μ_{jk} might be the predicted diversion ratio from some demand system such as a logit or random coefficients logit model. Among practitioners the most common assumption is that the diversion ratio is proportional to marketshare (which would be equivalent to demand following the plain IIA logit without covariates). An advantage of the shrinkage estimator is that it allows us to nest the plain IIA logit estimates of diversion currently used in practice and the experimental outcomes, depending on our choice of λ. When μ_{jt} is chosen as a function of the same observed dataset (including from estimated demand parameters) this is a form of an *Empirical Bayes* estimator. The development of Empirical Bayes shrinkage is attributed to Morris (1983) and has been widely used in applied microeconomics to shrink outliers from a distribution of fixed effects in teacher value added[10] or hospital quality.[11]

An estimator based on Assn 1 and 2 would be considered a Bayesian non-parametric estimator for the ATT because there is a corresponding parameter for each Δq_k observation in the dataset. We might want to incorporate the further assumption that the sum of estimated diversion ratios is equal to one (including diversion to "no purchase"), or that diversion is described by a (nonparametric) multinomial discrete-choice model. The straightforward multinomial generalization of Assn 2 would be the Dirichlet-Multinomial conjugacy. However, while the Dirichlet prior allows for each alternative to have a separate mean μ_{jk}, it only allows for a single precision parameter (or number of pseudo-observations) λ. To circumvent this we suggest a more flexible Bayesian non-parametric estimator for the ATT based on the over-parametrized normal, which allows for a separate mean and precision parameter for each alternative k and is a common technique in the statistics literature.[12]

Assumption 3. *"Unit Simplex":* $\sum_{\forall k} D_{jk} = 1$
$\eta_{jk}|\mu_{jk}, \sigma_{jk} \sim N(\mu_{jk}, \sigma_{jk})$, $\eta_{j0} = 0$, $D_{jk} = \frac{\exp[\eta_{jk}]}{\sum_{k'} \exp[\eta_{jk'}]}$
and $[\Delta q_0 \ldots \Delta q_K]|\Delta q_j, D_{j0} \ldots D_{jK} \sim Mult(\Delta q_j, D_{j0}, \ldots, D_{jK})$

Assn 3 implies that consumers who switch away from j now make a multinomial choice among a set of alternatives k including a "no purchase" or "outside option" denoted by $k = 0$. In some cases, this may be a stronger assumption than we are willing to make, because it requires data not just from the two merging parties, but from all of the parties in the industry. The advantage is that now we can use information on diversion from $j \to k'$ in order to learn about diversion from $j \to k$. This estimator also exhibits the same type of shrinkage as the beta-binomial model above. Many pseudo-observations imply a small σ_{jk} while fewer observations imply a larger σ_{jk} (and more shrinkage towards the prior mean). The "adding-up" constraint of multinomial choice may further shrink all probabilities towards zero or inflate them towards one in order to enforce the constraint implied by Assn 3. If there is no non-experimental variation in the choice set (the set of alternatives remains fixed)

[10]Chetty, Friedman, and Rockoff (2014) and Kane and Staiger (2008)
[11]Chandra, Finkelstein, Sacarny, and Syverson (2013)
[12]Gelman, Bois, and Jiang (1996) and Blei and Lafferty (2007).

then Assn 3 may not be so objectionable. However if options k and k' are similar to one another (red buses and blue buses) but never available at the same time, the multinomial assumption may not be reasonable, especially if we are interested in $D_{jk} + D_{jk'}$.

2.4 Identification Discussion

Despite the fact that $D_{jk} = \frac{\exp[\eta_{jk}]}{\sum_{k'} \exp[\eta_{jk'}]}$ superficially resembles a multinomial logit model, the estimator is in fact non-parametric. The rationale is that the data are only Δq_k, the difference in sales between the treatment and control. For each of the K observations in our data, we have two parameters (μ_{jk}, σ_{jk}), which leads statisticians to describe this as an *over-parametrized distribution*.

Though a multinomial logit model (or some more flexible variant) may look similar, identification is quite different. In those models, identification usually arises from observational variation in the set of alternatives, or observational variation in prices (through an instrument). We often use information such as how the price (or presence) of an alternative k' differentially affects demand for j and k in order to identify the parameters of the model. However, the goal in that case is the identification of a fully-specified model of demand that can be used to compute diversion (or conduct structural merger simulation) on any number of alternatives. For example, a logit model could be used to compute the diversion ratio of $D_{j'k}$ instead of D_{jk}, but with the caveat that it does not usually have enough parameters to exactly fit every possible $D_{j'k}$ that might be observed in the data.

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.

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*.[13] 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. If the relevant market is defined as all products sold in a snack-food vending machine, 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 was owned by at least seven firms between 1985 and 2001, including the Keebler Cookie Company (acquired by Kellogg in 2001), and the Presidential Baking Company (acquired by Keebler in 1998). *Zoo Animal Crackers* have a similarly complicated history, having been owned by Austin Quality Foods before they too were acquired by the Keebler Cookie Co. (which in turn was acquired by Kellogg).[14]

Our study measures diversion through the lens of a single medium-sized retail vending operator in the Chicago metropolitan area, Mark Vend Company. Each of Mark Vend's machines internally records price and quantity information. The data track total vends and

[13]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.

[14]Snack foods have an important historic role in market definition. A landmark case was brought by *Tastykake* in 1987 in an 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 a 2% marketshare nationwide, but a much larger share in the Northeast (including 50% of the New York market). Tastykake successfully argued that the relevant market was single-serving snack cakes rather than a broad category of snack foods involving cookies and candy bars. [Tasty Baking Co. v. Ralston Purina, Inc., 653 F. Supp. 1250 - Dist. Court, ED Pennsylvania 1987.]

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 a 38-month panel that runs from January 2006 to February 2009. 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 structural demand estimation. Very few "natural" stock-outs occur at our set of machines.[15] 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. Some products have very low levels of sales and we consolidate them with similar products within a category produced by the same manufacturer, until we are left with 73 'products' that form the basis of the rest of our exercise.[16]

In addition to the data from Mark Vend, we also collect data on the characteristics of each product online and through industry trade sources.[17] For each product, we note its manufacturer, as well as the following set of product characteristics: package size, number of servings, and nutritional information.[18]

3.1 Experimental Design

We ran four exogenous product removals with the help of Mark Vend Company. These represent a subset of a larger group of eight exogenous product removals that we have analyzed in two other projects, Conlon and Mortimer (2010) and Conlon and Mortimer (2015). Our experiment uses 66 snack machines located in professional office buildings and serviced by Mark Vend. 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 order to be able to run each product removal

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

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

[17]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.

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

over a shorter period of time across more machines.[19] These machines were also located on routes that were staffed by experienced drivers, which maximized the chance that the product removal would be successfully implemented. The 66 machines used for each treatment are 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 remove a product from all machines at a client site for a period of 2.5 to 3 weeks. The four products that we remove are the two best-selling products from either (a) chocolate maker Mars Incorporated (Snickers and Peanut M&Ms) or (b) cookie maker Kellogg's (Famous Amos Chocolate Chip Cookies and Zoo Animal Crackers). We refer to exogenously-removed products as the *focal products* throughout our analysis.[20]

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. Whenever a product was exogenously removed, 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, Mark Vend wanted to minimize the number of phone calls received in response to the stock-out events.

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 the data as they were collected. Drivers are generally paid a small commission on the

[19] Many high-volume machines are located in public areas (e.g., museums or hospitals), and feature demand patterns (and populations) that vary 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 have relatively stable demand patterns.

[20] Not reported here are two experiments on best-selling products from Pepsi's Frito Lay Division, which we omit for space considerations, and because Pepsi's products already dominate the salty snack category (which makes merger analysis less relevant). We also ran two additional experiments in which we removed two products at once; again we omit those for space considerations and because they don't speak to our diversion ratio example. These are analyzed in Conlon and Mortimer (2010) and Conlon and Mortimer (2015).

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. Technologically, of course, 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 "same-side" group of untreated machines. We chose this design, rather than randomly staggering the product removals, because we (and the participating clients) 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 client location hitting a busy period that temporarily induces long work hours and higher vending sales. Conversely, the design has the benefit that we can aggregate over all machines at a client site, and treat the entire site as if it were a single machine. Despite the imperfections of field experiments in general, these are often the kinds of tests 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

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 many retail environments. We often observe week-over-week sales that vary by over 20%, with 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. In our particular setting, many of the product removals were implemented during the summer of 2007, which was a high-point in demand at several sites, most likely due to macroeconomic conditions. In this case, using a simple control like previous weeks' sales, or overall average sales, can result in unreasonable treatment outcomes, such as overall sales increasing due to a product removal, or decreasing by more than the sales of the removed product.

The second challenge is that the data are recorded at the level of a service visit to a vending machine. It is 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 assume that sales are distributed uniformly among the business days in a service interval, and assign sales to weeks. 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.[21]

After converting all observations to a machine-week, we summarize the data in Table 2. Across our four treatments and 66 machines, we observe between 161-223 treated machine-weeks. In the untreated group, we observe 8,525 machine-weeks and more than 700,000 units sold. Each treatment week exposes around 2,700-3500 individuals, of which around 134-274 would have purchased the focal product in an average week. Each treatment lasts 2.5-3 weeks, which exposes roughly 400-1,600 individuals over the course of each treatment. This highlights one of the main challenges of measuring diversion experimentally: for the purposes of measuring the treatment effect, only individuals who would have purchased the focal product, had it been available, are considered "treated," yet we must expose many more individuals to the product removal, knowing that many of them were not interested in the focal product in the first place.

Throughout the analysis, our fundamental unit is a machine-week. For each machine-week in the treatment group we directly observe the sales of each product, and compare that to the average sales during the corresponding machine-weeks from the control group. As discussed in Section 2.3, we only include machine-weeks in the control group that come from the same machine, when the substitute product k was also available. This leads to a slightly different set of potential controls for each treatment. It also means that not every machine-week in the treatment group is included in the analysis for each substitute. The first four columns of Table 4 report, for each treatment and substitute product, the number of machine-weeks, the average number of controls, the change in substitute sales Δq_k, and the change in focal sales Δq_j. In columns (6)-(9), we restrict the set of machine-weeks in the control group to those that satisfy Assumption 1, and report the same quantities. For the Snickers removal,

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

if we examine substitution to Twix Caramel, we can see that there were 143 machine-weeks in our treated group, and on average there were 120.3 machine-weeks in the control group for each treated machine-week. Once we restrict to these control weeks (i.e. those for which sales did not increase and did not decrease by more than the expected Snickers sales), we are left with an average of 9.8 control machine-weeks per treatment machine-week. There are now 134 treatment machine-weeks instead of 143 treatment machine-weeks, because 9 treatment machine-weeks have no corresponding feasible control weeks. This is similar to the k-nearest neighbor matching approach of Abadie and Imbens (2006), except that we allow for a variable number of matches per treatment observation, and choose matches based on an economic feasibility criterion rather than a statistical distance criterion.

In an ideal experimental setting, the only variation in choice sets across machine-weeks would come from our experimental treatments. In that case, we would expect to see very little variation in the number of machine-weeks or feasible control observations across potential substitutes. For the Snickers experiment, the "outside good" option, reported in the last row, is always available. This indicates a maximum of 180 machine-weeks for our Snickers treatment, for which at least one control week satisfies Assumption 1. Only two products are available in all 180 machine-weeks (Dorito Nacho and Chocolate Chip Famous Amos), and several are available in nearly all of the machine-weeks (M&M Peanut, Rold Gold Pretzels, Zoo Animal Crackers, and Cheetos); but most products are available in fewer than half of the machine-week observations (and consequently in fewer than half of the 970 "treated" focal sales).

We begin with a non-parametric analysis of the experimental outcomes. For each treatment, we compute the average sales of each substitute product. We compare these average sales during treated periods to the distribution of control-week sales, and report the corresponding quantile of that distribution in Table 3. For example, when Snickers are removed, the average weekly sales of Twix exceed any weeks observed during the control period, and sales of Planters Peanuts correspond to the 96th percentile of the distribution of control week sales. Overall sales across machines during the treatment corresponds to the 74.4th percentile of overall control week sales. This provides some evidence in favor of restricting control observations on the basis of Assumption 1 in order to control for variation in overall demand.

4.1 Estimates of Diversion

The obvious estimator of the diversion ratio for each substitute product k and each focal (treated) product j is $D_{jk} = \frac{\Delta q_k}{\Delta q_j}$, which we can also interpret as the MLE from observing Δq_k successes in Δq_j binomial trials. The first panel of Table 5 reports this "raw" diversion ratio for the Snickers removal. For Twix, in the second row, $\Delta q_k = 289.6$ and $\Delta q_j = -702.4$ based on the 134 machine-weeks in which Twix was available. This implies a raw diversion ratio $D_{jk} = 41.2\%$. In the same table, we observe substitution from Snickers to Non-Chocolate Nestle products with only 3 machine-weeks in our sample.[22] This leads to $\Delta q_j = -10.5$ and $\Delta q_k = 9.4$ for an implied diversion ratio of $D_{jk} = 89.5\%$. Examining these raw diversion numbers may lead one to conclude that Non-Chocolate Nestle products are a closer substitute for Snickers than Twix. However, we observe more than 70 times as much information about substitution to Twix as we do to Non-Chocolate Nestle products.

Our best estimate of the diversion ratio for Snickers to Non-Chocolate Nestle products may actually be smaller than the diversion ratio of Snickers to Twix once we incorporate uncertainty into our estimates. Assumption 2 shrinks the diversion ratio towards the prior mean; when we have more treated individuals (and hence more information) we employ less shrinkage, and when we have fewer treated individuals, we employ more shrinkage. The beta prior requires $D_{jk} \in [0, 1]$. While the unadjusted estimates allow for $\Delta q_k < 0$ and $D_{jk} < 0$, under Assumption 2 we treat these values as if $\Delta q_k = 0$.

When we implement Assumption 2, we follow an Empirical Bayes procedure in choosing the prior. We assume that the prior mean of the diversion ratio for μ_{jk} is centered around the marketshare for product k, s_k. This is of some practical significance, because the assumption of logit demand, or diversion proportional to marketshare, is commonly used in practice by antitrust authorities.[23] Recall from Section 2.3 that λ tells us how much weight to put on our prior mean versus our experimental observations. The choice of λ allows us to nest the logit model and our experimental diversion measures in a single framework. When the experiment is less informative, we "shrink" our diversion estimates towards the logit.

We take the approach of choosing different numbers of pseudo-observations from the Beta prior distribution and examining robustness to the strength of the prior. Under the *strong prior* in Table 5 we assume that $N = 300$ pseudo-observations were observed before

[22] Non-Chocolate Nestle products include Willy Wonka candies such as Tart-N-Tinys, Chewy Tart-N-Tinys, Mix-ups, Mini Shockers, and Chewy Runts.

[23] Another alternative would be to use a more sophisticated and realistic model of consumer demand that relaxes the IIA property of the logit, such as the random coefficients logit model, and use the predicted diversion ratio from that model as our prior mean.

our experimental data. Under the *weak prior* we assume there is one pseudo-observation for each product in the choice set. Even with the relatively diffuse prior, we see that the estimate of diversion for Non-chocolate Nestle products shrinks from 89.5% to 12.4%, while the diversion estimate for Twix Caramel only shrinks from 41.2% to 37.9%. As expected, we see much more shrinkage when the number of treated individuals (Δq_j or change in focal sales) is small. We explore how estimated diversion ratios vary with alternative prior strengths in Table 6.

Assumption (3) adds the requirement that the diversion ratio of all substitute products (including substitution to the "outside good" option) sum to one. This embeds the idea that every individual not buying j must buy some other k or nothing at all. It eliminates some possibilities like substituting from j to the pair of products (k, k'), or that k and k' are substitutes. It also implicitly assumes that all substitute products are simultaneously available in the treatment and control. In order to construct a prior distribution, we now assume that $D_{jk} = \frac{\exp \eta_{jk}}{\sum_{k'} \exp \eta_{jk'}}$ where $\eta_{jk} \sim N(\mu_{jk}, \sigma^2_{jk})$. We follow the Empirical Bayes procedure, and use observed data from the control period to estimate $\mu_{jk} = \mu_k$ as the parameters of the plain IIA multinomial logit. Rather than fix the strength of the prior, we experiment with different values of $\sigma^2_k = \sigma^2$. In Table 7 we document how our estimates of diversion vary with σ^2 from $\sigma^2 = 0.25$ to $\sigma^2 = 100$.[24] We find that the strength of the prior has very little effect on the estimated diversion ratio, even when the prior is extremely diffuse ($\sigma^2 = 100$). This is different from the product-by-product binomial case in Table 6, where diversion estimates are relatively sensitive to the strength of the prior distribution. We observe shrinkage of the experimental estimates when there are fewer treated individuals (e.g., Non-Chocolate Nestle products have an estimated diversion ratio from Snickers of less than 1%) independent of the weight placed on the prior distribution. We observe relatively less shrinkage for products with more treated individuals (e.g., the estimated diversion from Snickers to Twix is 15.9%).

In Table 8, we report the posterior distribution of our diversion estimates under Assumption 3 and the very diffuse prior $\sigma^2 = 100$. We find that in most cases the posterior distribution defines a relatively tight 95% confidence interval, even when we have relatively few experimentally-treated individuals. On one hand this indicates our estimates are relatively precise and insensitive to the prior distribution. On the other, it demonstrates the power of Assumption 3 (the requirement that the sum of diversion ratios across all substitutes is equal to one). Essentially, this assumption alone, even with a very small effective

[24]A typical value of $\mu_k \approx -4$.

λ, pins down the posterior estimates of the diversion ratio. While the assumption appears relatively innocuous (most researchers are likely willing to assume a multinomial discrete choice framework) because it is so powerful in pinning down the diversion ratio estimates, we should be a little cautious. The important empirical content of Assumption 3 is determining what the appropriate set of products $\mathcal{K}$ is, such that $\sum_{k\in\mathcal{K}} D_{jk} = 1$. If, for example, we were interested in a merger where product j acquired both (k, k') but (k, k') were always rotated for one another and never available at the same time, we might want to vary the set of products over which we sum $D_{jk'}$ for each alternative: $\mathcal{K}_k$.[25]

Our finding that estimates of diversion that use Assumptions 1 and 2 can be sensitive to the prior, while those that use Assumptions 1 and 3 precisely pin down the diversion ratio, is of practical significance as well. One of the stated benefits of the unilateral effects approach is that it requires data only from the merging parties, and not from firms outside the merger.[26] The power of Assumption 3 indicates that measuring diversion to all substitute goods (rather than just k) can substantially improve our estimates of D_{jk}.[27] This suggests that although we need only (quasi)-experimental removals (or second-choice data) for the focal products involved in the merger, we should attempt to measure substitution to all available substitutes if possible.

4.2 Merger Evaluation

The goal behind estimating the diversion ratio is to allow regulators to perform prospective merger analysis. Within the unilateral-effects framework, the diversion ratio is the key input for calculating UPP or the GUPPI. These two measures also rely on estimates of price-cost margins. Under a typical structural approach to merger analysis, price cost margins are typically treated as unobserved, and researchers estimate them via a demand system and an assumption about firm conduct (such as static Bertrand-Nash pricing) [RE-WRITE]. In contrast, when the FTC or DOJ evaluate a merger, firms are compelled to provide measures of price-cost margins as part of the Hart-Scott-Rodino filings.

[25]Conlon and Mortimer (2013) show that assuming all products are always available introduces bias in structural parametric estimates of demand.

[26]The 2010 Horizontal Merger Guidelines include the phrase: *Diversion ratios between products sold by merging firms and those sold by non-merging firms have at most secondary predictive value.* We disagree with this statement in terms of statistical properties, rather than economic theory.

[27]In broad strokes, this phenomenon is well understood by statisticians. This is related to Stein's Paradox which shows that pooling information improves the parameter estimates for the mean of the multivariate normal, or the broader class of James-Stein shrinkage estimators. See Efron and Morris (1975) and James and Stein (1961).

The GUPPI for a merger in which the seller of product j acquired product k is $GUPPI_j = \frac{p_k - c_k}{p_j} \cdot D_{jk}$ (all evaluated at pre-merger values). An estimate of UPP would further require some information (or an assumption) about the potential size of the variable cost reduction from the proposed merger. When considering a merger between two manufacturers, the relevant margins are likely at the manufacturer level, rather than at the retail level, even though our measure of diversion is at the retail level. In our empirical example, we observe both the retail prices and prices paid by retailers to the manufacturers; what we do not observe are the marginal costs of the manufacturers. Thus, we would require additional assumptions in order to report UPP or the GUPPI. However, the vending machine context contains very little price variation at the retail level within a product category (all chocolate candy bars sell for 75 cents in our sample). Although there is some variation in wholesale prices, most are between 35 and 50 cents. Therefore, we don't expect margins to provide any additional information beyond the diversion ratio. This is probably true for many consumer products (e.g., the margin differences between blueberry and strawberry yogurt are likely insignificant); but in other contexts (such as automobiles or airlines) margin differences are likely an important part of the story.

Given our experimental estimates of D_{jk}, what precisely can we say about the likely effects of a merger? The first limitation is that if we have an estimate of D_{jk} we do not necessarily have an estimate of D_{kj}. Thus, while one may be able to make statements about how the acquisition of k is likely to affect the incentives to raise prices for j, one may not necessarily be able to make statements about how "being acquired by j" is likely to increase the price of k (unless we experimentally remove k as well). Assuming that diversion ratios are symmetric does not seem like a good idea, and places strong restrictions on the parametric form of demand.[28] A second issue is understanding what the relevant product or products are. If price-cost margins are the same, then if the owner of j acquires both k and k', we are interested in the sum of the diversion ratios: $D_{jk} + D_{jk'}$. Often mergers involve large numbers of brands.

An important remedy available to the antitrust agencies is to approve a merger conditional on some divestiture.[29] Define the diversion ratio from j to all products owned by a

[28]Consider the IIA logit model. In that model, diversion is proportional to market share. Thus the acquisition of a small brand by a large brand is unlikely to have much effect on the large brand, because the diversion ratio is small. It is more likely to have an effect on the small brand, because many consumers would be predicted to divert to the large brand.

[29]Some recent examples include gate slots at specific airports for the American/USAirways merger, or requiring that the entire Modelo business within the US be spun-off and divested during its acquisition by Anheuser-Busch InBev.

target firm $\mathcal{F}$ as: $D_{j,\mathcal{F}} = \sum_{k \in \mathcal{F}} D_{j,k}$, and then consider removing one or more products from the set of products involved in the merger. The idea is to divest a few closely-competing products from a proposed merger, so that the aggregated diversion ratio $D_{j,\mathcal{F}}$ is below some critical threshold (often 5% or 10%).

[RE-ORDER THIS AND PREVIOUS PARAGRAPHS] Our empirical example in Table 9 follows exactly this pattern. We consider the potential price effects of a merger only on the products that we experimentally removed by examining the sum of diversion ratios from these products to all of the products of potential targets $\mathcal{F}$. We then propose a divestiture of a key substitute product controlled by the target $\mathcal{F}$, and recompute the diversion ratio to all of the target's products absent the divested product. In some circumstances, divestiture of one or two key products might alleviate price concerns around a particular merger.

When we examine a potential acquisition where Kellogg's (Pop-Tarts, Zoo Animal Crackers, Famous Amos Cookies, Cheez-it, Rice Krispie Treats) acquires Kraft's snack food division (Oreos, Lorna Doone, Planters Peanuts, Cheese Nips, and other Nabisco products). We can look at the effects on both of Kellogg's major products (Zoo Animal Crackers and Famous Amos) of the Kraft acquisition. We find that the diversion ratio from Zoo Animal Crackers to all Kraft products is 5.80%, and the diversion of Chocolate Chip Famous Amos is 11.85%. Both of these are above the 5% threshold, and thus would need to demonstrate substantial cost synergies to justify the merger. However, if Kraft were to divest its Planter's Peanut line, the diversion ratios drop to 3.36% and 3.09% respectively. This suggests a potential remedy that might allow the antitrust authorities to drop opposition to the merger.

Likewise one could consider an acquisition by Mars (Snickers, M&M's, Milky Way, Three Musketeers, Skittles) of Nestle's US confections business (Butterfinger, Raisinets, assorted Willy Wonka fruit flavored candies). If one is worried about the price effects that the acquisition might have on Snickers or M&M Peanut (the two largest brands in the chocolate category) we find that diversion from those products to Nestle products is 5.71% and 6.30% respectively, but that if Nestle were to divest Butterfinger, the diversion ratios would drop to 1.26% and 4.51% respectively. Again, this might be enough to convince the antitrust authorities not to block a proposed acquisition.[30]

Our hope is that these examples highlight both the advantages of our approach (that it is easy to detect which mergers require further investigation and which divestitures to consider), but also some of the limitations. For example, we can look at the effect on Snickers or M&M

[30]While this kind of divestiture may sound less realistic than Kraft divesting the Planters Peanuts line, these kinds of agreements are actually commonplace in the confections industry. For example in the United States, Kit-Kat is a Hershey product, but outside the United States Kit-Kat is a Nestle product.

Peanut that the acquisition of Butterfinger might have, but we cannot say anything about the likely effect on the prices of Butterfinger of a Mars acquisition without conducting that experiment as well. This suggests that we either need observational/quasi-experimental data on many different stockout events, or we need some *ex ante* idea of which products are likely to have larger price impacts of the merger in order to tailor our experiments. The second limitation, which is not a limitation of our approach but of the unilateral effects approach more generally, is that it ignores diversion to existing brands. In the Snickers experiment, more than half of consumers already substitute to another Mars product, yet this has no bearing on the analysis of a proposed merger with Nestle (though it might if we considered the price effects on Butterfinger). This highlights what is likely to be a more general pattern in the unilateral effects approach: when large brands acquire smaller brands, the likely concern is the price of the smaller brand.

5 Conclusion

The 2010 revision to the Horizontal Merger Guidelines de-emphasized market definition and traditional concentration measures such as HHI in favor of a unilateral effects approach based on UPP or GUPPI. This unilateral effects approach holds prices, costs, and competitive responses fixed, and the key input is the diversion ratio, which measures how closely two products substitute for one another.

We show that the diversion ratio can be interpreted as the marginal treatment effect of an experiment in which the price of one product is increased by a small amount. An important characteristic of many retail settings is that category-level sales can be more variable than product-level market shares. In practice, this makes most experiments that consider small price changes under-powered. We also show that second-choice data arising from randomized experiments, quasi-experiments (such as stockouts), or second-choice survey data, can be used to estimate an average diversion ratio, where the average is taken over all possible prices from the pre-merger price to the choke price. We derive conditions based on economic primitives such as the curvature of demand, whereby the average diversion ratio from second-choice data (ATE) is a good approximation for the MTE.

We conduct randomized field experiments, where we exogenously remove products from consumers' choice sets and measure the ATE directly. We provide a set of three relatively minimal assumptions, derived from consumer theory, which allow for relatively precise estimates of the diversion ratio even in noisy environments: (1) that a product removal cannot increase sales, nor decrease sales by more than the expected sales of the removed product

(2) that the diversion ratio is defined on the unit interval and has some prior distribution and (3) that the diversion ratio for all possible substitutes is defined on the unit simplex. We find benefits from measuring diversion not only between products involved in a proposed merger, but also from merging products to non-merging products.

We develop a simple method to recover the diversion ratio from data, which enables us to combine both experimental and quasi-experimental measures with structural estimates as prior information. An Empirical Bayes shrinkage approach enables us to use prior information (or potentially structural estimates) when experimental measures are not available, or when they are imprecisely measured, and to rely on experimental measures when they are readily available. This facilitates the combination of both first- and second-choice consumer data. We show that these approaches are complements rather than substitutes. Structural demand estimates rely either on variation in the prices of choices across contexts or markets, or on the set of choices that are available to consumers in order to identify parameters. These estimators struggle when there is little variation in the data. In contrast, randomized controlled trials work best when there is little to no variation in the availability or prices of products, except for the variation induced by the treatment.

Our hope is that this makes a well-developed set of quasi-experimental and treatment effects tools available both to researchers in industrial organization and also to antitrust practitioners. While the diversion ratio can be obtained experimentally, doing so is not trivial, and researchers should think carefully about (1) which treatment effect their experiment (or quasi-experiment) is actually identifying; and (2) what the identifying assumptions required for estimating a diversion ratio implicitly assume about the structure of demand.

References

Abadie, A., and G. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74(1), 235–267.

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.

Angrist, J. D., K. Graddy, and G. W. Imbens (2000): "The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish," *The Review of Economic Studies*, 67(3), 499–527.

Armstrong, T. (2013): "Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply," Working Paper.

Berry, S., J. Levinsohn, and A. Pakes (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112(1), 68–105.

Blei, D. M., and J. D. Lafferty (2007): "A correlated topic model of science," *The Annals of Applied Statistics*, pp. 17–35.

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

Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2013): "Healthcare Exceptionalism? Productivity and Allocation in the US Healthcare Sector," Discussion paper, National Bureau of Economic Research.

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

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104(9), 2593–2632.

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

Conlon, C. (2011): "The MPEC Approach to Empirical Likelihood Estimation of Demand," Unpublished Mansucript. Columbia University.

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

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

——— (2015): "All Units Discount: Experimental Evidence from the Vending Industry," Working Paper.

Efron, B., and C. Morris (1975): "Data analysis using Stein's estimator and its generalizations," *Journal of the American Statistical Association*, 70(350), 311–319.

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.

Gelman, A., F. Bois, and J. Jiang (1996): "Physiological pharmacokinetic analysis using population modeling and informative prior distributions," *Journal of the American Statistical Association*, 91(436), 1400–1412.

Hastings, J. S. (2004): "Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California," *American Economic Review*, 94(1), 317–328.

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.

Heckman, J. J., and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

Jaffe, S., and E. Weyl (2013): "The first-order approach to merger analysis," *American Economic Journal: Microeconomics*, 5(4), 188–218.

James, W., and C. Stein (1961): "Estimation with quadratic loss," in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 361–379.

Judd, K. L., and B. Skrainka (2011): "High performance quadrature rules: how numerical integration affects a popular model of product differentiation," CeMMAP working papers CWP03/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Kane, T. J., and D. O. Staiger (2008): "Estimating teacher impacts on student achievement: An experimental evaluation," Discussion paper, National Bureau of Economic Research.

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.

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.

Morris, C. N. (1983): "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78(381), 47–55.

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.

Shapiro, C. (1995): "Mergers with differentiated products," *Antitrust*, 10, 23.

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.

Team, S. (2015): "RStan: the R interface to Stan, Version 2.8.0," .

Werden, G. J. (1996): "A Robust Test for Consumer Welfare Enhancing Mergers among Sellers of Differentiated Products," *Journal of Industrial Economics*, 44(4), 409–13.

Werden, G. J., and L. Froeb (2006): "Unilateral competitive effects of horizontal mergers," *Handbook of AntitrustEconomics*.

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.

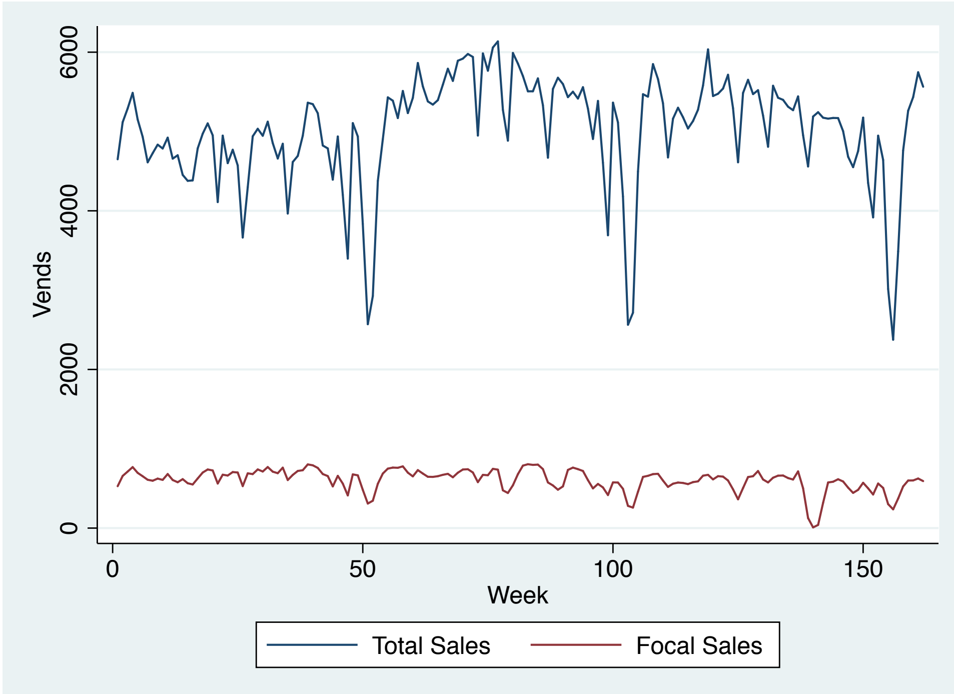

Figure 1: Total Overall Sales and Sales of Snickers and M&M Peanuts by Week

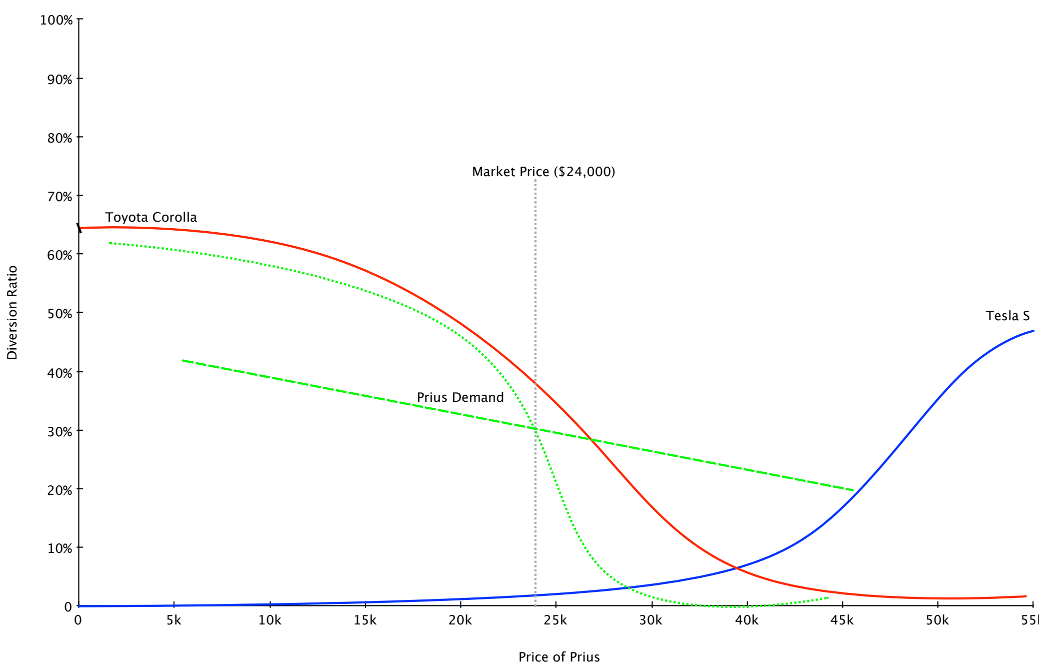

Figure 2: Diversion from Prius (Hypothetical Example)

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

Table 1: Manufacturer Market Shares and HHI's by Category and Total

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)

	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

Table 2: Summary Statistics

† Numbers for Snickers removal. Summary statistics for other removals differ minimally because of different definition of the starting day of the week.

* Focal sales during the control period. Focal sales during the treatment are close to zero. Any deviation from zero occurs because of the apportionment of service visit level sales to weekly sales.

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	Zoo Animal Cracker Austin	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
Famous Amos Cookie Removal				
Mars	M&M Peanut	309.7	319.5	46.4
Mars	Snickers	301.2	316.6	52.0
Pepsi	Rold Gold (Con)	158.9	285.3	80.0
Pepsi	Cheeto LSS	248.7	260.7	64.8
Snyders	Snyders (Con)	210.1	236.4	52.8
Pepsi	Sun Chip LSS	150.2	225.5	100.0
Pepsi	Ruffles (Con)	206.9	218.3	62.4
	Total	4890.2	5262.4	64.0
M&M Peanut Removal				
Mars	Snickers	300.9	411.8	99.2
Snyders	Snyders (Con)	209.7	279.0	76.8
Pepsi	Rold Gold (Con)	158.9	276.9	80.8
Pepsi	Cheeto LSS	248.6	251.0	47.2
Mars	Twix Caramel	167.9	213.8	90.4
Kellogg	Zoo Animal Cracker Austin	182.6	198.0	65.6
Pepsi	Baked Chips (Con)	169.4	194.7	68.0
	Total	4886.1	5315.5	65.6

Table 3: Quantile of Average Treatment Period Sales in the Empirical Distribution of Control Period Sales.

Control Mean is the average number of sales of a given product (or all products) over all control weeks. Treatment Mean is the average number of sales of a given product (or all products) over all treatment weeks. A treatment week is any week in which at least one machine was treated. For client sites that were not treated during these weeks (because treatment occurs at slightly different dates at different sites), we use the average weekly sales for the client site when it was under treatment (otherwise we would be comparing treatment weeks with different number of treated machines in them). Treatment Quantile indicates in which quantile of the distribution of control-week sales the treatment mean places.

		Unrestricted Control Machine-Weeks					Restricted Control Machine-Weeks				
Manufacturer	Product	Trt'd Mach-Weeks	Avg # Controls Per Trt	Δ Subst Sales	Δ Focal Sales	Raw Diversion	Trt'd Mach-Weeks	Avg # Controls Per Trt	Δ Subst Sales	Δ Focal Sales	Raw Diversion
Nestle	Nonchoc Nestle (Con)	6	80.3	14.1	-19.8	71.1	3	8.7	9.4	-10.5	89.5
Mars	M&M Peanut	186	120.3	482.4	-915.9	52.7	176	10.0	375.5	-954.3	39.4
Mars	Twix Caramel	143	120.3	339.6	-682.6	49.7	134	9.8	289.6	-702.4	41.2
Misc	Farleys (Con)	22	40.9	41.0	-121.2	33.8	18	4.6	14.9	-114.2	13.0
Hershey	Choc Herhsey (Con)	51	51.9	62.1	-210.0	29.6	41	8.8	29.8	-179.6	16.6
Mars	M&M Milk Chocolate	104	116.1	114.7	-454.6	25.2	97	10.6	71.8	-457.4	15.7
Pepsi	Rold Gold (Con)	186	82.8	215.5	-874.6	24.6	174	7.6	161.4	-900.1	17.9
Nestle	Butterfinger	63	95.5	78.8	-355.7	22.1	61	7.9	72.9	-362.8	20.1
Kraft	Planters (Con)	143	94.8	154.8	-708.0	21.9	136	7.9	78.0	-759.9	10.3
Kellogg	Rice Krispies Treats	20	93.8	15.9	-72.9	21.8	17	6.5	17.7	-66.5	26.7
Mars	Choc Mars (Con)	12	67.5	5.2	-34.7	14.9	11	16.2	6.4	-32.7	19.7
Hershey	Payday	2	84.0	1.4	-9.7	14.4	2	8.5	1.1	-9.8	10.9
Kellogg	Zoo Animal Cracker Austin	187	120.3	132.0	-923.6	14.3	177	9.5	65.7	-970.2	6.8
Kellogg	Choc SandFamous Amos	74	113.4	52.7	-369.9	14.2	69	10.0	33.9	-404.2	8.4
Hershey	Sour Patch Kids	34	124.9	17.0	-134.3	12.6	33	12.7	10.8	-152.9	7.1
Kellogg	Brown Sug Pop-Tarts	6	74.7	3.6	-30.4	11.8	6	8.2	2.3	-33.1	7.0
Pepsi	Sun Chip LSS	166	117.8	91.7	-814.5	11.3	159	9.1	45.3	-866.1	5.2
Sherwood	Ruger Wafer (Con)	162	82.7	80.9	-734.5	11.0	151	7.6	24.5	-778.0	3.1
Nestle	Choc Nestle (Con)	1	21.0	0.9	-9.3	9.2	0				
Kar's Nuts	Kar Sweet&Salty Mix 2oz	113	116.6	50.1	-565.7	8.8	104	8.9	27.6	-597.1	4.6
Kellogg	Choc Chip Famous Amos	190	119.0	81.8	-932.9	8.8	180	10.0	44.8	-971.8	4.6
Kraft	Fig Newton	6	77.0	2.1	-29.6	7.2	6	5.8	0.6	-31.3	2.0
Nestle	Raisinets	143	121.7	47.6	-678.8	7.0	133	10.0	11.6	-697.3	1.7
Pepsi	FritoLay (Con)	113	94.9	32.7	-507.0	6.4	104	9.7	16.8	-515.7	3.3
Pepsi	Baked Chips (Con)	176	113.5	49.5	-883.5	5.6	166	10.1	33.5	-911.7	3.7
Misc	Farleys Mixed Fruit Snacks	137	93.3	34.9	-666.8	5.2	129	7.2	13.0	-686.5	1.9
Pepsi	Dorito Blazin Buffalo Ranch LSS	95	57.6	20.0	-494.0	4.0	87	5.2	-27.6	-503.1	-5.5
Mars	Combos (Con)	132	78.2	27.5	-682.6	4.0	119	6.6	7.6	-663.6	1.2
Kellogg	Cheez-It Original SS	159	119.6	25.3	-794.1	3.2	150	10.4	2.1	-819.9	0.3
Mars	Starburst Original	31	108.5	4.2	-138.7	3.0	29	11.6	-1.7	-137.6	-1.2
Pepsi	Cheeto LSS	187	120.3	27.0	-918.7	2.9	177	10.0	-46.2	-957.4	-4.8
Mars	Marathon Chewy Peanut	7	83.0	0.9	-42.0	2.1	6	6.5	-5.0	-50.4	-9.9
Misc	BroKan (Con)	3	43.0	0.0	-0.2	1.5	3	42.0	0.0	0.0	
Kraft	Cherry Fruit Snacks	71	123.1	5.3	-398.1	1.3	68	9.3	-5.3	-419.3	-1.3
Misc	Popcorn (Con)	77	113.9	1.5	-387.1	0.4	76	9.8	-19.8	-425.2	-4.6
Snyders	Snyders (Con)	145	104.7	0.6	-630.6	0.1	137	9.2	-76.6	-668.6	-11.5
Misc	Rasbry Knotts	147	109.4	-1.8	-736.1	-0.2	136	9.3	-4.5	-727.7	-0.6
Pepsi	Ruffles (Con)	156	124.4	-2.9	-774.1	-0.4	148	10.4	-42.2	-794.9	-5.3
Kraft	Lorna Doone Shortbread Cookies	43	123.6	-0.8	-197.8	-0.4	41	11.3	-4.6	-202.3	-2.3
Misc	Other Pastry (Con)	4	91.0	-0.1	-17.0	-0.5	3	8.7	-0.1	-12.8	-0.6
Pepsi	Quaker Strwbry Oat Bar	44	78.2	-1.3	-186.6	-0.7	39	9.6	-7.3	-174.0	-4.2

		Unrestricted Control Machine-Weeks					Restricted Control Machine-Weeks				
Manufacturer	Product	Trt'd Mach-Weeks	Avg # Controls Per Trt	Δ Subst Sales	Δ Focal Sales	Raw Diversion	Trt'd Mach-Weeks	Avg # Controls Per Trt	Δ Subst Sales	Δ Focal Sales	Raw Diversion
Kellogg	Strwbry Pop-Tarts	162	118.1	-6.0	-792.7	-0.8	154	9.9	-40.5	-819.4	-4.9
General Mills	Nature Valley Swt&Salty Alm	49	107.0	-2.3	-214.8	-1.1	43	9.6	-42.4	-195.3	-21.7
Pepsi	Chs PB Frito Cracker	48	95.0	-2.7	-220.5	-1.2	45	9.0	-6.4	-227.9	-2.8
Kraft	Ritz Bits Chs Vend	74	127.4	-5.3	-404.9	-1.3	71	9.4	0.2	-424.0	0.0
Mars	Nonchoc Mars (Con)	35	108.1	-2.1	-154.3	-1.3	31	13.1	1.0	-134.8	0.7
Kar's Nuts	KarNuts (Con)	40	99.3	-2.6	-183.8	-1.4	35	8.0	-27.7	-188.4	-14.7
Kraft	100 Cal Chse Nips Crisps	20	93.8	-1.1	-72.9	-1.5	17	6.5	-6.3	-66.5	-9.4
Pepsi	Smartfood LSS	67	125.5	-7.8	-365.3	-2.1	65	9.2	-25.0	-388.2	-6.4
Kellogg	Cherry Pop-Tarts	28	87.9	-3.0	-125.4	-2.4	28	7.5	2.4	-155.4	1.6
Mars	Milky Way	11	94.8	-1.4	-42.4	-3.3	9	4.6	-0.5	-37.9	-1.4
Pepsi	Dorito Nacho LSS	190	119.7	-37.2	-928.3	-4.0	180	10.0	-57.9	-969.1	-6.0
Misc	Hostess Pastry	16	114.4	-3.2	-76.6	-4.1	15	15.9	-11.7	-78.7	-14.8
Pepsi	Cheetos Flaming Hot LSS	69	124.8	-15.4	-371.5	-4.1	66	9.1	-22.3	-372.9	-6.0
Pepsi	Grandmas Choc Chip	119	114.6	-29.9	-589.7	-5.1	111	9.8	-36.3	-580.7	-6.3
Kraft	100 Cal Oreo Thin Crisps	23	94.0	-4.2	-75.3	-5.6	20	11.9	1.2	-66.5	1.7
Mars	Skittles Original	132	122.9	-37.8	-650.9	-5.8	125	9.7	-49.0	-672.5	-7.3
Misc	Cliff (Con)	4	32.0	-1.6	-22.9	-6.9	4	3.0	-1.6	-24.7	-6.6
Snyders	Jays (Con)	161	98.0	-58.3	-775.8	-7.5	150	8.6	-87.8	-809.4	-10.8
Pepsi	Frito LSS	154	106.0	-69.5	-749.8	-9.3	144	9.4	-84.4	-798.1	-10.6
General Mills	Oat n Honey Granola Bar	37	118.2	-24.9	-204.4	-12.2	36	9.0	-29.7	-197.1	-15.1
Misc	Salty Other (Con)	31	115.3	-18.8	-147.3	-12.8	30	12.5	-11.9	-163.8	-7.3
Pepsi	Lays Potato Chips 1oz SS	155	64.9	-96.2	-713.7	-13.5	143	5.5	-112.5	-744.1	-15.1
Misc	Salty United (Con)	11	76.5	-6.0	-30.1	-20.0	9	16.7	-9.6	-26.1	-36.8
Mars	3-Musketeers	3	52.0	-2.9	-8.3	-35.4	2	11.0	0.0	0.0	
Hershey	Twizzlers	55	53.9	-83.4	-216.4	-38.5	46	7.8	-75.6	-192.8	-39.2
	Outside Good	190	120.5	-982.6	-929.3	-105.7	180	10.0	460.9	-970.2	47.5

Table 4: Removal 1 (Snickers Removal) Raw Diversion Ratios - Unrestricted and Restricted Control Machine-Weeks

Assumption 1 (removing a product cannot increase total sales during a period, and cannot decrease total sales by more than the expected sales of the removed product) is imposed when selecting Restricted Control Machine-Weeks, but is not imposed when selecting Unrestricted Control Machine Weeks. Both restricted and unrestricted machine-weeks require that the focal product and the substitute of interest are available in the control machine-week, and the substitute is available in the treatment machine-week.

Trt'd Mach-Weeks shows the number of treated machine-weeks for which there was at least one control machine-week. Avg # Controls Per Trt is the average number of control machine-weeks per treatment machine-week over all treatment machine-weeks. Δ Subst Sales shows the change in substitute product sales from the control to the treatment period, while Δ Focal Sales shows the analogous change for focal product sales. Raw Diversion is the ratio of the change in substitute product sales to the absolute value of the change in focal product sales.

Manufacturer	Product	Treated Machine Weeks	Δ Subst Sales	Δ Focal Sales	Raw Diversion	Beta-Binomial (Weak Prior) Diversion	Beta-Binomial (Strong Prior) Diversion	Multinomial Diversion
Snickers Removal								
Nestle	Nonchoc Nestle (Con)	3	9.4	-10.5	89.5	12.4	3.1	0.7
Mars	Twix Caramel	134	289.6	-702.4	41.2	37.9	29.5	15.9
Mars	M&M Peanut	176	375.5	-954.3	39.4	37.0	30.8	18.4
Kellogg	Rice Krispies Treats	17	17.7	-66.5	26.7	13.5	5.0	1.3
Nestle	Butterfinger	61	72.9	-362.8	20.1	17.1	11.2	4.5
Mars	Choc Mars (Con)	11	6.4	-32.7	19.7	6.5	2.0	0.4
Pepsi	Rold Gold (Con)	174	161.4	-900.1	17.9	16.8	13.9	7.5
Hershey	Choc Herhsey (Con)	41	29.8	-179.6	16.6	12.2	6.3	2.0
Mars	M&M Milk Chocolate	97	71.8	-457.4	15.7	13.8	9.8	4.1
Misc	Farleys (Con)	18	14.9	-114.2	13.0	8.3	3.7	1.0
Hershey	Payday	2	1.1	-9.8	10.9	1.4	0.4	0.1
Kraft	Planters (Con)	136	78.0	-759.9	10.3	9.6	7.8	3.8
	Outside Good	180	460.9	-970.2	47.5			23.1
Zoo Animal Crackers Removal								
Hershey	Payday	2	0.4	-0.4	84.7	0.6	0.1	
Kellogg	Rice Krispies Treats	13	23.5	-37.8	62.2	23.2	7.2	3.0
Misc	Salty United (Con)	6	10.4	-18.9	55.1	12.6	3.4	1.2
Kraft	100 Cal Oreo Thin Crisps	13	14.9	-37.8	39.4	14.7	4.5	1.8
Pepsi	Rold Gold (Con)	132	114.4	-440.8	25.9	22.9	16.2	9.9
Hershey	Choc Herhsey (Con)	30	33.6	-132.6	25.3	17.1	7.9	3.8
Misc	Hostess Pastry	11	14.7	-62.2	23.7	11.8	4.4	1.8
Kraft	100 Cal Chse Nips Crisps	13	8.7	-37.8	23.1	8.6	2.6	1.1
Mars	Milky Way	9	7.0	-30.8	22.6	7.5	2.2	0.9
Mars	Snickers	145	92.4	-483.6	19.1	17.3	13.0	7.6
Mars	M&M Peanut	142	77.7	-469.4	16.6	15.0	11.4	6.5
Mars	Twix Caramel	110	50.2	-339.0	14.8	12.7	8.7	4.6
	Outside Good	145	240.5	-482.9	49.8			22.0
Famous Amos Cookie Removal								
Nestle	Choc Nestle (Con)	1	0.8	-0.2	300.0	1.2	0.3	
Hershey	Choc Herhsey (Con)	38	48.6	-66.8	72.7	36.9	13.4	7.2
Kraft	100 Cal Oreo Thin Crisps	29	20.7	-43.3	47.9	19.2	6.1	3.1
Pepsi	Sun Chip LSS	139	143.6	-355.7	40.4	34.4	22.7	15.8
Hershey	Payday	2	2.6	6.8	38.9			
Misc	Salty United (Con)	18	9.9	-28.7	34.6	10.7	3.1	1.5
Pepsi	Chs PB Frito Cracker	34	26.9	-83.6	32.1	18.2	7.1	3.7
Kraft	Planters (Con)	121	82.1	-332.6	24.7	20.9	13.7	8.8
Kellogg	Choc SandFamous Amos	57	28.0	-122.0	22.9	15.1	6.8	3.7
Mars	Milky Way	26	13.9	-71.6	19.5	10.3	3.9	1.9
Pepsi	Dorito Blazin Buffalo Ranch LSS	72	38.1	-224.2	17.0	13.3	7.5	4.4
Pepsi	Frito LSS	119	49.9	-313.2	15.9	13.4	8.9	5.3
	Outside Good	156	192.9	-399.1	48.3			20.9

Manufacturer	Product	Treated Machine Weeks	Δ Subst Sales	Δ Focal Sales	Raw Diversion	Beta-Binomial (Weak Prior) Diversion	Beta-Binomial (Strong Prior) Diversion	Multinomial Diversion
				M&M Peanut Removal				
Misc	Hostess Pastry	11	12.5	-38.6	32.5	12.3	4.0	1.8
Mars	Snickers	218	296.6	-1239.3	23.9	22.9	19.9	16.5
Kellogg	Brown Sug Pop-Tarts	10	10.0	-43.5	22.9	9.2	2.9	1.4
Misc	Cliff (Con)	1	0.4	-1.8	22.2	0.6	0.1	0.0
Nestle	Nonchoc Nestle (Con)	1	0.9	-4.6	19.5	1.3	0.3	0.1
Mars	M&M Milk Chocolate	99	73.5	-529.6	13.9	12.5	9.2	6.3
Mars	Twix Caramel	176	110.9	-1014.3	10.9	10.4	8.9	6.8
Kellogg	Rice Krispies Treats	46	22.4	-220.2	10.2	7.9	4.4	2.5
Hershey	Twizzlers	62	33.0	-333.0	9.9	8.3	5.3	3.3
Hershey	Choc Herhsey (Con)	32	15.7	-160.0	9.8	7.0	3.5	1.9
Kellogg	Cherry Pop-Tarts	25	12.5	-160.3	7.8	5.6	2.8	1.6
Mars	Nonchoc Mars (Con)	45	14.6	-201.3	7.3	5.5	3.0	1.7
	Outside Good	218	606.2	-1238.5	48.9			36.4

Table 5: Raw and Bayesian Diversion Ratios.

Treated Machine Weeks shows the number of treated machine-weeks for which there was at least one control machine-week. Δ Subst Sales shows the change in substitute product sales from the control to the treatment period, while Δ Focal Sales shows the analogous change for focal product sales. Raw Diversion is the ratio of the change in substitute product sales to the absolute value of the change in focal product sales.

Beta-Binomial (Weak Prior) Diversion and Beta-Binomial (Strong Prior) Diversion are diversion ratios calculated under Assumptions 1 (Substitutes) and Assumption 2 (Unit Interval). The weak prior uses the number of products in the choice set during the treatment period, which varies from 64 [56 if include the double removal] to 66, as the number of pseudo-observations. The strong prior uses 300 pseudo-observations. Multinomial Diversion is the diversion ratio calculated under Assumptions 1 (Substitutes) and Assumption 3 (Unit Simplex).

The products included in this table are the 12 products with highest raw diversion ratio.

Manuf	Product	Δ Focal Sales	No Prior	Beta-Bin Diversion $n = J^{\dagger}$	Beta-Bin Diversion $n = 150$	Beta-Bin Diversion $n = 300$	Beta-Bin Diversion $n = 600$
			Snickers Removal				
Nestle	Nonchoc Nestle (Con)	-10.5	89.5	12.4	5.9	3.1	1.6
Mars	Twix Caramel	-702.4	41.2	37.9	34.3	29.5	23.2
Mars	M&M Peanut	-954.3	39.4	37.0	34.5	30.8	25.5
Kellogg	Rice Krispies Treats	-66.5	26.7	13.5	8.4	5.0	2.9
Nestle	Butterfinger	-362.8	20.1	17.1	14.3	11.2	7.8
Mars	Choc Mars (Con)	-32.7	19.7	6.5	3.5	2.0	1.0
Pepsi	Rold Gold (Con)	-900.1	17.9	16.8	15.7	13.9	11.6
Hershey	Choc Herhsey (Con)	-179.6	16.6	12.2	9.1	6.3	3.9
			Zoo Animal Crackers Removal				
Hershey	Payday	-0.4	84.7	0.6	0.3	0.1	0.1
Kellogg	Rice Krispies Treats	-37.8	62.2	23.2	12.7	7.2	3.9
Misc	Salty United (Con)	-18.9	55.1	12.6	6.3	3.4	1.8
Kraft	100 Cal Oreo Thin Crisps	-37.8	39.4	14.7	8.0	4.5	2.4
Pepsi	Rold Gold (Con)	-440.8	25.9	22.9	19.8	16.2	12.1
Hershey	Choc Herhsey (Con)	-132.6	25.3	17.1	12.0	7.9	4.7
Misc	Hostess Pastry	-62.2	23.7	11.8	7.2	4.4	2.5
Kraft	100 Cal Chse Nips Crisps	-37.8	23.1	8.6	4.7	2.6	1.4
			Famous Amos Cookie Removal				
Nestle	Choc Nestle (Con)	-0.2	300.0	1.2	0.6	0.3	0.2
Hershey	Choc Herhsey (Con)	-66.8	72.7	36.9	22.5	13.4	7.4
Kraft	100 Cal Oreo Thin Crisps	-43.3	47.9	19.2	10.8	6.1	3.3
Pepsi	Sun Chip LSS	-355.7	40.4	34.4	28.9	22.7	16.1
Hershey	Payday	6.8	38.9				
Misc	Salty United (Con)	-28.7	34.6	10.7	5.6	3.1	1.7
Pepsi	Chs PB Frito Cracker	-83.6	32.1	18.2	11.6	7.1	4.1
Kraft	Planters (Con)	-332.6	24.7	20.9	17.5	13.7	9.8
			M&M Peanut Removal				
Misc	Hostess Pastry	-38.6	32.5	12.3	6.9	4.0	2.3
Mars	Snickers	-1239.3	23.9	22.9	21.7	19.9	17.2
Kellogg	Brown Sug Pop-Tarts	-43.5	22.9	9.2	5.2	2.9	1.6
Misc	Cliff (Con)	-1.8	22.2	0.6	0.3	0.1	0.1
Nestle	Nonchoc Nestle (Con)	-4.6	19.5	1.3	0.6	0.3	0.2
Mars	M&M Milk Chocolate	-529.6	13.9	12.5	11.0	9.2	7.0
Mars	Twix Caramel	-1014.3	10.9	10.4	9.8	8.9	7.6
Kellogg	Rice Krispies Treats	-220.2	10.2	7.9	6.1	4.4	2.9

Table 6: Sensitivity of Beta-Binomial Diversion to Number of Pseudo Observations

† Number of pseudo observations is the number of products in the choice set during treatment period - 66, 64, 65, and 65, respectively.

Δ Focal Sales shows the change in focal product sales from the control to the treatment period. No Prior is the raw diversion calculated as the ratio of the change in substitute product sales to the absolute value of the change in focal product sales.

Beta-Bin Diversion is the diversion ratio calculated under Assumptions 1 (Substitutes) and Assumption 2 (Unit Interval), using different number of pseudo-observations.

The products included in this table are the 8 products with highest raw diversion ratio.

Manuf	Product	Δ Focal Sales	No Prior	Multinom Diversion $\sigma^2 = 0.25$	Multinom Diversion $\sigma^2 = 1$	Multinom Diversion $\sigma^2 = 10$	Multinom Diversion $\sigma^2 = 100$
			Snickers Removal				
Nestle	Nonchoc Nestle (Con)	-10.54	89.45	0.59	0.61	0.66	0.67
Mars	Twix Caramel	-702.39	41.23	15.00	15.63	15.87	15.90
Mars	M&M Peanut	-954.30	39.35	17.58	18.16	18.39	18.41
Kellogg	Rice Krispies Treats	-66.45	26.68	1.05	1.18	1.28	1.30
Nestle	Butterfinger	-362.82	20.11	3.89	4.27	4.43	4.45
Mars	Choc Mars (Con)	-32.73	19.70	0.45	0.40	0.43	0.44
Pepsi	Rold Gold (Con)	-900.11	17.93	6.98	7.37	7.52	7.54
	Outside Good	-970.22	47.50	22.17	22.82	23.07	23.10
			Zoo Animal Crackers Removal				
Kellogg	Rice Krispies Treats	-37.80	62.22	2.38	2.75	2.95	2.98
Misc	Salty United (Con)	-18.91	55.09	1.07	1.13	1.23	1.25
Kraft	100 Cal Oreo Thin Crisps	-37.80	39.40	1.49	1.68	1.81	1.84
Pepsi	Rold Gold (Con)	-440.80	25.95	8.95	9.61	9.87	9.90
Hershey	Choc Herhsey (Con)	-132.57	25.34	3.13	3.56	3.77	3.80
Misc	Hostess Pastry	-62.15	23.66	1.45	1.63	1.77	1.79
Kraft	100 Cal Chse Nips Crisps	-37.80	23.12	0.97	1.00	1.07	1.09
	Outside Good	-482.91	49.81	20.59	21.56	21.91	21.97
			Famous Amos Cookie Removal				
Hershey	Choc Herhsey (Con)	-66.84	72.72	5.90	6.76	7.14	7.19
Kraft	100 Cal Oreo Thin Crisps	-43.29	47.89	2.33	2.77	3.02	3.06
Pepsi	Sun Chip LSS	-355.68	40.37	14.45	15.36	15.72	15.78
Misc	Salty United (Con)	-28.69	34.57	1.15	1.28	1.43	1.46
Pepsi	Chs PB Frito Cracker	-83.65	32.13	2.92	3.43	3.70	3.73
Kraft	Planters (Con)	-332.61	24.69	7.72	8.43	8.71	8.76
Kellogg	Choc SandFamous Amos	-122.04	22.91	2.91	3.39	3.63	3.68
	Outside Good	-399.12	48.33	19.53	20.53	20.89	20.94
			M&M Peanut Removal				
Misc	Hostess Pastry	-38.58	32.52	1.26	1.58	1.80	1.85
Mars	Snickers	-1239.29	23.93	15.63	16.23	16.46	16.47
Kellogg	Brown Sug Pop-Tarts	-43.53	22.91	0.98	1.19	1.36	1.40
Misc	Cliff (Con)	-1.80	22.22	0.30	0.13	0.03	0.01
Nestle	Nonchoc Nestle (Con)	-4.56	19.51	0.35	0.20	0.14	0.14
Mars	M&M Milk Chocolate	-529.58	13.87	5.37	5.99	6.25	6.28
Mars	Twix Caramel	-1014.32	10.94	6.08	6.55	6.73	6.76
	Outside Good	-1238.49	48.95	35.25	36.05	36.32	36.37

Table 7: Sensitivity of Multinomial Posterior Mean to Variance of Prior

Δ Focal Sales shows the change in focal product sales from the control to the treatment period. No Prior is the raw diversion calculated as the ratio of the change in substitute product sales to the absolute value of the change in focal product sales.
Multinom Diversion is the diversion ratio calculated under Assumptions 1 (Substitutes) and Assumption 3 (Unit Simplex), using priors of different strength.
The products included in this table are the 7 products with highest raw diversion ratio.

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	0.30	0.51	0.65	0.81	1.17
Mars	Twix Caramel	15.90	14.31	15.33	15.89	16.46	17.55
Mars	M&M Peanut	18.41	16.82	17.85	18.40	18.96	20.05
Kellogg	Rice Krispies Treats	1.30	0.78	1.09	1.27	1.48	1.94
Nestle	Butterfinger	4.45	3.51	4.10	4.44	4.78	5.49
Mars	Choc Mars (Con)	0.44	0.16	0.31	0.41	0.54	0.85
Pepsi	Rold Gold (Con)	7.54	6.46	7.16	7.54	7.92	8.67
	Outside Good	23.10	21.34	22.48	23.09	23.72	24.91
Zoo Animal Crackers Removal							
Kellogg	Rice Krispies Treats	2.98	1.92	2.56	2.94	3.37	4.27
Misc	Salty United (Con)	1.25	0.59	0.97	1.20	1.48	2.14
Kraft	100 Cal Oreo Thin Crisps	1.84	1.04	1.51	1.80	2.13	2.88
Pepsi	Rold Gold (Con)	9.90	8.25	9.30	9.88	10.48	11.67
Hershey	Choc Herhsey (Con)	3.80	2.63	3.36	3.76	4.22	5.14
Misc	Hostess Pastry	1.79	1.00	1.46	1.74	2.07	2.80
Kraft	100 Cal Chse Nips Crisps	1.09	0.50	0.83	1.05	1.31	1.91
	Outside Good	21.97	19.61	21.13	21.96	22.79	24.41
Famous Amos Cookie Removal							
Hershey	Choc Herhsey (Con)	7.19	5.38	6.51	7.15	7.82	9.22
Kraft	100 Cal Oreo Thin Crisps	3.06	1.93	2.59	3.02	3.48	4.49
Pepsi	Sun Chip LSS	15.78	13.51	14.96	15.75	16.56	18.19
Misc	Salty United (Con)	1.46	0.71	1.13	1.42	1.73	2.48
Pepsi	Chs PB Frito Cracker	3.73	2.46	3.24	3.69	4.18	5.21
Kraft	Planters (Con)	8.76	7.04	8.12	8.73	9.36	10.64
Kellogg	Choc SandFamous Amos	3.68	2.48	3.21	3.64	4.11	5.10
	Outside Good	20.94	18.45	20.04	20.92	21.79	23.56
M&M Peanut Removal							
Misc	Hostess Pastry	1.85	0.99	1.49	1.80	2.16	2.93
Mars	Snickers	16.47	14.84	15.89	16.46	17.05	18.17
Kellogg	Brown Sug Pop-Tarts	1.40	0.67	1.08	1.36	1.67	2.36
Misc	Cliff (Con)	0.01	0.00	0.00	0.00	0.00	0.08
Nestle	Nonchoc Nestle (Con)	0.14	0.00	0.04	0.10	0.20	0.51
Mars	M&M Milk Chocolate	6.28	4.99	5.80	6.26	6.74	7.71
Mars	Twix Caramel	6.76	5.61	6.34	6.75	7.16	7.99
	Outside Good	36.37	34.20	35.63	36.37	37.12	38.56

Table 8: Posterior Distribution of Multinomial Diversion with $\sigma^2 = 100$

The products included in this table are the 7 products with highest raw diversion ratio.

Proposed Merger	Diversion Direction	Diversion Ratio	Proposed Divestiture	Diversion Ratio Under Divestiture
Mars & Hershey	Snickers to Hershey	2.83	Reese's Peanut Butter Cups	2.83*
	M&M Peanut to Hershey	7.14	Reese's Peanut Butter Cups	5.30
Mars & Kraft	Snickers to Kraft	3.97	Planters Peanuts	0.16
	M&M Peanut to Kraft	4.21	Planters Peanuts	0.62
Mars & Nestle	Snickers to Nestle	5.71	Butterfinger	1.26
	M&M Peanut to Nestle	6.30	Butterfinger	4.51
Mars & Kellogg's	Snickers to Kellogg's	8.53	Famous Amos Cookies†	4.58
	M&M Peanut to Kellogg's	5.64	Famous Amos Cookies†	5.48
	Zoo Animal Crackers to Mars	21.80	Famous Amos Cookies†	21.80
Kellogg's & Kraft	Zoo Animal Crackers to Kraft	5.80	Planters Peanuts	3.36
	Choc Chip Famous Amos to Kraft	11.85	Planters Peanuts	3.09

Table 9: Hypothetical Mergers with Forced Divestitures

* Reese's Peanut Butter Cups are unavailable in all treatment weeks for this experiment.
† Divestiture of both "Choc Chip Famous Amos" and "Choc SandFamous Amos".

A Appendix:

A.1 Diversion 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:

$$D_{jk} = \frac{\Delta Q_k}{\Delta Q_j} = \frac{\beta_{kj}\Delta p_j}{\beta_{jj}\Delta p_j} = \frac{\beta_{kj}}{\beta_{jj}} \tag{10}$$

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:

$$\frac{\Delta \log(Q_k)}{\Delta \log(Q_j)} \approx \underbrace{\frac{\Delta Q_k}{\Delta Q_j}}_{D_{jk}} \cdot \frac{Q_j(\mathbf{p})}{Q_k(\mathbf{p})} \quad = \quad \frac{\varepsilon_{kj}\Delta \log(p_j)}{\varepsilon_{jj}\Delta \log(p_j)} = \frac{\varepsilon_{kj}}{\varepsilon_{jj}}$$

$$D_{jk} \quad \approx \quad \frac{Q_k(\mathbf{p})}{Q_j(\mathbf{p})} \cdot \frac{\varepsilon_{kj}}{\varepsilon_{jj}} \tag{11}$$

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:

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

$$S_{jt} = \frac{\exp[\tilde{v}_{jt}]}{1 + \sum_{k \in a_t} \exp[\tilde{v}_{kt}]} = \frac{V_{jt}}{IV(a_t)}$$

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

$$\widehat{D_{jk}} = \frac{\frac{\partial S_k}{\partial p_j}}{\left|\frac{\partial S_j}{\partial p_j}\right|} = \frac{\alpha S_k S_j}{\alpha S_j (1 - S_j)} = \frac{S_k}{(1 - S_j)}$$

$$\overline{D_{jk}} = \frac{\frac{e^{V_k}}{1 + \sum_{l \in a \setminus j} e^{V_l}} - \frac{e^{V_k}}{1 + \sum_{l' \in a} e^{V'_l}}}{0 - \frac{e^{V_j}}{1 + \sum_{l \in a} e^{V_l}}} = \frac{S_k}{(1 - S_j)}$$

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: $D_{jk} = \frac{\partial^2 q_k}{\partial p_j^2} / \frac{\partial^2 q_j}{\partial p_j^2}$.

$$\frac{\partial^2 q_j}{\partial p_j^2} = \alpha^2 (1 - 2S_j)(S_j - S_j^2)$$

$$\frac{\partial^2 q_k}{\partial p_j^2} = -\alpha^2 (1 - 2S_j) S_j S_k$$

$$\frac{\frac{\partial^2 q_k}{\partial p_j^2}}{\frac{\partial^2 q_j}{\partial p_j^2}} = \frac{S_k}{1 - S_j} = D_{jk}$$

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 $\frac{\partial V_{ik}}{\partial p_j} = \alpha$,

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

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

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. We generate a single market with J products, and compute the $J \times J$ matrix of diversion ratios two ways. The MTE method is by computing $\frac{\partial q_k}{\partial p_j} / |\frac{\partial q_j}{\partial p_j}|$

A.2 Discrepancy Between Average and Marginal Treatment Effects

We can perform a Monte Carlo study to analyze the extent to which the average treatment effect deviates from the marginal treatment effect. We generate data by simulating from a random coefficients logit model with a single random coefficient on price. Our simulations follow the procedure in Armstrong (2013), Judd and Skrainka (2011) and Conlon (2011) where prices are endogenously solved for via a Bertrand-Nash game given the other utility parameters, rather than directly drawn from some distribution.

We generate the data in the following manner: $u_{it} = \beta_0 + x_j\beta_1 - \alpha_i p_j + \xi_j + \varepsilon_{ij}$ and $mc_j = \gamma_0 + \gamma_1 x_j + \gamma_2 z_j + \eta_j$ where $x_j, z_j \sim N(0,1)$, with $\xi_j = \rho\omega_{j1} + (1-\rho)\omega_{j2} - 1$ and

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mu	0.1	0.1	1	1	2	3	3
Sigma	0.5	1	0.5	1	1	1	2
Outside Good Share	0.97	0.85	0.91	0.77	0.94	0.99	0.90
Avg Own Elas	-5.37	-3.50	-4.64	-3.12	-3.70	-4.41	-1.93
Avg Max Discrepancy	1.51	3.68	1.72	2.37	2.13	2.04	3.00
Std. Max Discrepancy	0.36	1.18	0.50	0.66	0.60	0.59	1.32
Worst Case Avg ATE	7.14	14.18	9.29	12.38	10.80	9.08	15.01
Worst Case Avg MTE	5.62	10.50	7.58	10.01	8.67	7.04	12.01

Table 10: Simulation comparing ATE and MTE for Random Coefficients Logit

α	-0.500	-0.500	-0.500	-0.500	-1.000	-1.000	-1.000	-1.000
σ_p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
σ_x	0.500	1.000	2.000	3.000	0.500	1.000	2.000	3.000
s_0	0.031	0.037	0.063	0.100	0.109	0.123	0.167	0.213
own elas	-0.318	-0.318	-0.319	-0.310	-1.523	-1.520	-1.514	-1.511
avg max dev	0.028	0.107	0.340	0.636	0.038	0.130	0.386	0.630
std max dev	0.023	0.084	0.270	0.501	0.028	0.089	0.260	0.461
max dev ATE	15.287	15.863	16.684	18.942	13.102	13.646	15.307	17.444
max dev MTE	15.260	15.757	16.344	18.306	13.064	13.516	14.921	16.814
pct dev	0.167	0.630	1.924	3.207	0.273	0.915	2.486	3.522

Table 11: Monte Carlo Simulations

$\eta_j = \rho\omega_{j1} + (1-\rho)\omega_{j3} - 1$ and $(\omega_1, \omega_2, \omega_3) \sim^{i.i.d.} U[0,1]$. Following Armstrong (2013) and Conlon (2011), we use the values $\beta = [-3, 6]$ and $\gamma = [2, 1, 1]$ and $\rho = 0.9$. To mimic our empirical example we let there be $J = 30$ products and assign each product at random to one of 5 firms. We solve for prices in a Bertrand-Nash equilibrium.

For each of our sets of trials, we let $\alpha_i \sim -lognormal(\mu, \sigma)$ and we vary the values of price heterogeneity in the population by changing (μ, σ). We simulate 100 trials from each (μ, σ) pair are report characteristics of that market (average outside good share, average own price elasticity) as well as describe the discrepancy between the ATE and the MTE approach to computing diversion. We report those results for the pair of products in each trial with the largest discrepancy between the ATE and MTE calculations.

Though there are some simulations where the $ATE < MTE$, in the vast majority of simulations the random coefficients model with a lognormally distributed price coefficient implies that using the stock-out based ATE overstates the true MTE for the diversion ratio by 1-3 points in the worst-case scenario (the maximum over the entire $J \times J$ matrix of diversion ratios). The degree of overstatement appears to be decreasing in the lognormal location parameter (as consumers become more price sensitive) and increasing in the dispersion parameter (as consumers become more heterogeneous).

α	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
σ_p	0.250	0.250	0.250	0.250	0.500	0.500	0.500	0.500
σ_x	0.500	1.000	2.000	3.000	0.500	1.000	2.000	3.000
s_0	0.116	0.130	0.171	0.217	0.134	0.149	0.182	0.223
own elas	-1.510	-1.514	-1.487	-1.479	-1.476	-1.459	-1.431	-1.399
avg max dev	0.624	0.670	0.842	1.033	2.538	2.578	2.398	2.447
std max dev	0.227	0.253	0.376	0.565	0.878	0.971	0.836	0.987
max dev ATE	12.479	12.821	14.785	16.757	13.019	13.328	13.572	16.109
max dev MTE	11.858	12.154	13.943	15.727	10.490	10.782	11.255	13.920
pct dev	5.594	5.778	6.179	6.617	26.241	25.913	23.570	19.643

Table 12: Monte Carlo Simulations

α	-2.000	-2.000	-2.000	-2.000	-2.000	-2.000	-2.000	-2.000
σ_p	0.250	0.250	0.250	0.250	0.500	0.500	0.500	0.500
σ_x	0.500	1.000	2.000	3.000	0.500	1.000	2.000	3.000
s_0	0.591	0.580	0.565	0.553	0.578	0.575	0.562	0.555
own elas	-3.846	-3.836	-3.832	-3.834	-3.479	-3.487	-3.498	-3.523
avg max dev	0.351	0.443	0.771	1.245	1.164	1.177	1.373	1.807
std max dev	0.096	0.147	0.367	0.836	0.324	0.354	0.474	0.784
max dev ATE	8.349	8.950	11.787	15.819	9.522	10.096	12.414	15.775
max dev MTE	7.998	8.507	11.016	14.574	8.358	8.919	11.041	13.968
pct dev	4.640	5.463	7.097	8.395	14.941	14.131	13.127	13.416

Table 13: Monte Carlo Simulations

α	-4.000	-4.000	-4.000	-4.000	-4.000	-4.000	-4.000	-4.000
σ_p	0.500	0.500	0.500	0.500	1.000	1.000	1.000	1.000
σ_x	0.500	1.000	2.000	3.000	0.500	1.000	2.000	3.000
s_0	0.989	0.987	0.975	0.951	0.965	0.962	0.944	0.928
own elas	-7.735	-7.718	-7.811	-7.895	-5.183	-5.253	-5.440	-5.548
avg max dev	0.151	0.236	0.859	1.010	2.001	1.961	1.576	4.352
std max dev	0.059	0.087	0.340	0.851	0.587	0.634	0.717	2.611
max dev ATE	1.185	2.090	6.923	10.958	6.968	7.354	8.882	22.866
max dev MTE	1.034	1.854	6.064	9.973	4.967	5.393	7.344	20.779
pct dev	15.557	13.608	15.475	12.682	44.081	40.072	25.468	25.107

Table 14: Monte Carlo Simulations

A.3 Stan Code for MCMC Estimator

This is code for the R library *stan* (Team 2015) which recovers the MCMC estimator of the diversion ratio under assumptions (1)-(3).

```
data {
    int J;               // number of products, including outside good
    int N[J];            // number of trials
    int y[J];            // number of successes for each product j
    real mu_prior[J];    // mean of the distribution of alpha
    real sigma_prior[J]; // standard deviation of the distribution of alpha
}

parameters {
    row_vector[J] alpha;       // probability of success = exp(alpha[j])/(sum(exp(alpha[j])))
}

transformed parameters {
    row_vector[J] theta;
    for (j in 1:J)
        theta[j] <- exp(alpha[j])/(sum(exp(alpha))); // don't normalize the outside good
}

model {
    for (j in 1:J)
    alpha[j] ~ normal(mu_prior[j], sigma_prior[j]);

    for (j in 1:J) {
    y[j] ~ binomial(N[j], theta[j]);
    }
}
```