Empirical Properties of Diversion Ratios *

Christopher T. Conlon†
Julie Holland Mortimer‡

July 4, 2018

Abstract

A diversion ratio, which measures the fraction of consumers that switch from one product to an alternative after a price increase, is a central calculation of interest to antitrust authorities for analyzing horizontal mergers. Two ways to measure diversion are: the ratio of estimated cross-price to own-price demand derivatives, and second-choice data. Policy-makers may be interested in either, depending on whether they are concerned about the potential for small but widespread price increases, or product discontinuations. We estimate diversion in two applications – using observational price variation and experimental second-choice data respectively – to illustrate the trade-offs between different empirical approaches. Using our estimates of diversion, we identify candidate products for divestiture in a hypothetical merger.

*An earlier version of this paper was circulated under the title *An Experimental Approach to Merger Analysis.* We thank Mark Stein, Bill Hannon, and the drivers at Mark Vend Company for implementing the experiment used in this paper, providing data, and generally educating us about the vending industry. We thank Dennis Carlton, Joe Farrell, Ken Hendricks, Aviv Nevo, Dan O'Brien, Bill Rogerson, Ralph Winter, and seminar participants at Columbia University, the University of Western Ontario, the FTC Microeconomics Conference, 'I.O. Fest' at UC Berkeley, the Dartmouth Winter IO Conference, and the antitrust conference at the University of British Columbia for comments. Andrew Copland, Bogdan Genchev, Tom Gole, and Joonwoo Park 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

A diversion ratio, which measures the fraction of consumers that switch from one product to an alternative after a price increase, is one of the best ways economists have for understanding the nature of competition between sellers. Diversion ratios can be understood through the lens of a Nash-in-prices equilibrium when sellers offer differentiated products. Two products with a high degree of differentiation face lower diversion and softer price competition, whereas two products with a high degree of similarity to competing goods face higher diversion and tougher price competition.

Not surprisingly, diversion ratios are a central calculation of interest to antitrust authorities for analyzing horizontal mergers. The current U.S. merger guidelines, released in 2010, place greater weight on diversion ratios relative to concentration measures more commonly used to understand competition in settings with homogeneous goods (e.g., the Herfindahl-Hirschman Index (HHI)).[1] In the context of merger reviews, antitrust authorities identify the concept of 'unilateral effects' as being important for understanding the impact of a proposed merger. Unilateral effects of a merger arise when competition between the products of the merged firm is reduced because the merged firm internalizes substitution between its jointly-owned products.[2] This can lead to an increase in the price of the products of the merged firm, potentially harming consumers. Diversion ratios are the key statistic of interest for measuring unilateral effects. The current U.S. merger guidelines, released in 2010, note:

> *Diversion ratios between products sold by one merging firm and products sold by the other merging firm can be very informative for assessing unilateral price effects, with higher diversion ratios indicating a greater likelihood of such effects.*

Thus, holding competitive responses fixed, antitrust agencies will be more concerned about mergers that involve products with higher diversion ratios, because the scope for price increases due to unilateral effects is thought to be greater.

Although the use of diversion ratios in antitrust policy is well understood theoretically, in practice, one needs to estimate diversion ratios. The U.S. Guidelines discuss diversion ratios as being calculated from an estimated demand system, or observed from consumer survey

[1] Researchers have pointed out a number of concerns with using concentration measures or other functions of market share to capture the strength of competition. One concern is that such measures require one to define a market; another is that they do not capture the closeness of competition when products are differentiated, as most products are thought to be.

[2] In contrast, the concept of harm via 'coordinated effects' arises if a proposed merger increases the probability that firms in the industry will be able to successfully coordinate their behavior in an anti-competitive way.

data or in a firm's course of business. In this paper, we analyze different ways of estimating diversion ratios and characterize their empirical properties.

The researcher or antitrust authority may prefer different measurements of diversion in different settings. For example, if the antitrust authority is concerned with the potential for small but widespread price increases, they may want to evaluate diversion by analyzing estimated own- and cross-price derivatives at pre-merger prices. In contrast, if the antitrust authority is concerned with the potential for product discontinuations, second-choice data may be more informative. To clarify this point, we interpret a diversion ratio as a treatment effect of an experiment in which the treatment is "not purchasing product j." The treated group consists of consumers who would have purchased j at pre-existing prices, but no longer purchase j at a higher price. The diversion ratio measures the outcome of the treatment, (i.e., the fraction of consumers who switch from j to a substitute product k).

When policy-makers are interested in measuring the effect of treating only those consumers who substitute away from j after a very small price increase, they are implicitly evaluating a marginal treatment effect (MTE) at pre-merger prices.[3] A challenge of directly implementing such an experiment is that treating a small number of the most price-sensitive individuals may lack 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). 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 coincides with the evaluation of the MTE at pre-merger prices. 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 random-coefficients logit or log-linear models, do not feature constant diversion, and the ATE may diverge from the MTE evaluated at pre-merger prices.

A related question for the antitrust authority is whether one can reliably estimate diversion ratios using data from only the merging entities.To consider this question, it's useful to consider two concepts: an *aggregate diversion ratio*, which Katz and Shapiro (2003) define as the "percentage of the total sales lost by a product when its price rises that are captured by all of the other products in the candidate market," and a *diversion matrix*, which we define as a matrix whose off-diagonal elements report diversion between each pair of products that could potentially be considered for inclusion in a market, and whose diagonal elements report

[3] As we discuss later, one can view treatment effects estimators for price increases of different sizes as local average treatment effects (LATE).

diversion to the outside good.[4] Discrete choice models of demand imply a "summing up" constraint so that each row of the diversion matrix (i.e., aggregate diversion plus diversion to the outside good) sums to one.

We consider the empirical properties of diversion ratios in two applications. In the first application, we estimate two discrete-choice models of demand.We use data from Nevo (2000) to explore the properties of three different measures of diversion: a MTE evaluated at pre-merger prices, using a random-coefficients logit model; an ATE, estimated by simulating a product removal in the same random-coefficients model; and a 'plain vanilla' logit model, which assumes constant diversion proportional to marketshares. We show that the ATE and MTE measures differ by 6% on average for each product's closest substitute, and by about 8.3% across all substitutes. Substitution can be both over- and understated. A 'plain vanilla' logit model that assumes constant diversion substantially understates substitution to the best substitute, and overstates substitution to the outside good compared to the ATE or MTE measures.

In the second application, we construct an empirical estimator for the ATE measure of the diversion ratio by exogenously removing individual products from a local market in a large-scale experiment. Specifically, we remove products from a set of vending machines and track subsequent substitution patterns. The experimental setting precludes us from estimating diversion that would be relevant to a small price change because we are not able to exogenously change prices, but it does not require any parametric restrictions or any restrictions on aggregate diversion. In order to control for unobserved demand shocks, we select valid controls and impose a simple requirement that product removals cannot increase total sales, nor decrease total sales by more than the sales of the product removed. Having matched to these control observations, we consider, in turn, two additional assumptions about economic primitives and examine how they help to estimate experimental measures of the diversion ratio. The first assumption is that diversion to any single product is between 0 and 100 percent. We incorporate this assumption through a non-parametric Bayesian shrinkage estimator. We find that this improves our estimates of diversion, although our estimates are sensitive to the strength of the prior. Next, we impose the assumption that aggregate diversion plus diversion to the outside good sums to one. Our Bayesian shrinkage estimator incorporates this assumption by nesting the parametric structural estimates of diversion and the (quasi)-experimental measures in a single framework. With the "summing

[4] Thus, aggregate diversion represents the sum of the off-diagonal elements along each row of a diversion matrix. When defining aggregate diversion, Katz and Shapiro (2003) also impose the condition that "The aggregate diversion ratio must lie between zero and 100 percent."

up" constraint, even a very weak prior yields precise estimates of diversion ratios.

Our results highlight two important points: (1) Observing data from all products within the market, rather than only products involved in a merger, is important when estimating diversion ratios; and (2) in discrete-choice demand systems, the "summing up" constraint may play a more important role for identification than the parametric distribution of error terms. Our applications also illustrate the fact that different measures of diversion may be relevant to policy-makers in different settings. Several recent merger cases have been concerned with the potential for small but widespread price increases, such as in airline prices, and consumer goods and services.[5] Other cases have centered around the potential for product discontinuations, such as in hospital and airline networks, and in several business-to-business markets.[6]

Finally, our empirical exercise demonstrates how two different measures of diversion can be obtained in practice (i.e., through demand estimation or exogenous product removals), how different measures of diversion might vary, and how to design and conduct experiments to measure diversion. Using the estimates of diversion from our second application, we consider a hypothetical merger between Mars and Kellogg. We analyze diversion from key products of each firm to the brands of the other firm. The exercise illustrates the ability of diversion estimates to identify candidate products for divestiture requirements.

1.1 Related Literature

A second goal of the paper is to bring together two literatures – the applied theoretical literature that motivates the use of diversion for understanding merger impacts, and an applied econometric literature that articulates estimation challenges in settings for which the treatment effect of a policy can vary across individuals and may be measured with error.

By exploring the assumptions required for a credible (quasi)-experimental method of measuring diversion, we connect directly to the theoretical literature discussing the use and measurement of the diversion ratio.[7] Farrell and Shapiro (2010) suggest that firms themselves may track diversion in their 'normal course of business,' or that diversion ratios may be uncovered in Hart-Scott-Rodino filings. Hausman (2010) argues that the only acceptable

[5] Examples include the 2008 acquisition of Anheuser-Busch by InBev, and the American-US Air merger in 2015 (Das 2017).

[6] Examples include the discontinuation of some data storage products in the 2016 Dell-EMC merger, and route consolidation in the 2008 Delta-Northwest merger (Josephs 2018).

[7] 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).

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. A different strand of the applied theoretical 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. We don't take a stand on this question.[8]

In spirit, our approach is 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. That approach does not extend to a differentiated products setting because the requisite monotonicity condition may no longer be satisfied. Our second empirical application illustrates a differentiated products setting in which the average diversion ratio is identified from second-choice data alone, even though the separate own- and cross-price elasticities may not be. This highlights the economic content of (even partial) second-choice data, which have been found to be valuable in the structural literature on demand estimation (Berry, Levinsohn, and Pakes 2004).[9]

The paper proceeds as follows. Section 2 describes the theoretical framework behind the use of diversion ratios and discusses alternative ways to measure diversion using a treatment effects framework. Our first empirical application is provided in section 3, which estimates a discrete-choice model of demand that implicitly assumes that aggregate diversion plus diversion to the outside good sums to one. Section 4 describes our second application, which analyzes an experimental setting in the snack foods industry. Section 5 develops two

[8]This literature goes back to at least Shapiro (1995) and Werden (1996), and is well summarized in reviews by Farrell and Shapiro (2010) and Werden and Froeb (2006). A debate about the relationship between measures of upward pricing pressure (UPP) and merger simulations has developed since the release of the 2010 Horizontal Merger Guidelines including: Carlton (2010), Schmalensee (2009), Willig (2011), and Hausman (2010). In related work, Jaffe and Weyl (2013) incorporate an estimated pass-through rate to map anticipated opportunity cost effects of a merger into price effects. Miller, Remer, Ryan, and Sheu (2016) and Cheung (2011) find that the price effects of a merger, and errors in predicting these effects, depend on the nature of competition among non-merging firms, and whether prices are strategic substitutes or strategic complements. Miller and Weinberg (2017) explores the possibility that the merger between Anheuser-Busch Inbev and SABMiller facilitated tacit collusion through coordinated effects.

[9]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. Nevo and Whinston (2010) respond by pointing out that, while retrospective merger analysis is valuable, the salient 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 when both the proposal and approval of mergers is endogenous. This relates to a much older debate going back to Leamer (1983), and discussed more recently by Heckman (2010), Leamer (2010), Keane (2010), Sims (2010), Stock (2010), and Einav and Levin (2010).

estimators for an ATE measure of diversion outside of the context of discrete-choice demand models. In 6, we present results using different measures of diversion, discuss the role that aggregate diversion plays in our estimates, and consider the impacts of divestiture under several hypothetical mergers through the lens of our estimated diversion measures. Section 7 concludes.

2 Theoretical Framework

The rationale for focusing on diversion ratios to understand the potential impact of a merger comes from an underlying supply-side model in which firms produce differentiated goods and compete according to a Nash-in-prices equilibrium. Farrell and Shapiro (2010) present such a model to motivate the key constructs of the 2010 U.S. merger guidelines, and we review their results here.[10]

For simplicity, consider a single market composed of J single-product firms, where firm j sets the price of product j to maximize profits: $\pi_j = (p_j - c_j) \cdot q_j$. When firms (j,k) merge, firm j now considers the profits of firm k when setting her price. We can examine how this affects the first order condition for p_j holding all other prices p_{-j} fixed:

$$\begin{aligned}
\arg\max_{p_j} \quad & (p_j - c_j) \cdot q_j(p_j, p_{-j}) + (p_k - c_k) \cdot q_k(p_j, p_{-j}) \\
0 \quad &= \quad q_j + (p_j - c_j) \cdot \frac{\partial q_j}{\partial p_j} + (p_k - c_k) \cdot \frac{\partial q_k}{\partial p_j} \\
p_j \quad &= \quad -q_j / \frac{\partial q_j}{\partial p_j} + c_j + (p_k - c_k) \cdot \underbrace{\frac{\partial q_k}{\partial p_j} / - \frac{\partial q_j}{\partial p_j}}_{D_{jk}}
\end{aligned}$$

Now we can use the definition of the diversion ratio $D_{jk}(p_j, p_{-j}) = \frac{\partial q_k}{\partial p_j} / |\frac{\partial q_j}{\partial p_j}|$ and we can re-write firm j's best response in terms of the own price elasticity $\epsilon_{jj}(p_j, p_{-j})$:

$$p_j(p_{-j}) = \underbrace{\frac{\epsilon_{jj}}{\epsilon_{jj}+1}}_{\text{Lerner Markup}} \underbrace{\left[c_j \cdot \underbrace{(1 - e_j)}_{\text{efficiency}} + \underbrace{(p_k - c_k) \cdot D_{jk}(p_j, p_{-j})}_{\text{opp cost}} \right]}_{UPP} \tag{1}$$

Equation (1) shows us how post-merger prices are set. The usual elasticity-based markup is applied to marginal costs, with two additions from the merger. First, the merger may create

[10] We use slightly different notation to aid in our empirical applications later.

some efficiencies e_j that reduce costs, and second, the merger increases the opportunity cost of selling j, as some consumers leaving good j are now recaptured by product k. This opportunity cost has two inputs: the gross margin for product k: $(p_k - c_k)$ and the diversion ratio, D_{jk}, which measures the fraction of consumers leaving j who switch to k as p_j rises (holding all other prices p_{-j} fixed). The term in brackets is referred to as "Upward Pricing Pressure," or UPP. It is worth noting that equation (1) denotes only the best response for $p_j(p_{-j})$, and is not an equilibrium price.

2.1 A Matrix of Diversion Ratios

We can define a $J \times J$ matrix of diversion ratios where the (j,k)-th element is $D_{jk}(\mathbf{p})$. Rather than report $D_{jj} = -1$, we report D_{j0} (the diversion to the outside good) along the diagonal.

$$D(\mathbf{p}) = \begin{bmatrix} D_{10} & D_{12} & D_{13} \\ D_{21} & D_{20} & D_{23} \\ D_{31} & D_{32} & D_{30} \end{bmatrix}$$

This matrix is useful to make a number of conceptual points: (a) if all products are substitutes (rather than complements) and consumers make discrete choices, then each row of the matrix must sum to one: $D(\mathbf{p}) \times \mathbf{1_J} = \mathbf{1_J}$; (b) the sum of the off-diagonal elements along a row j is known as *aggregate diversion* for product j;[11] (c) most parametric models of demand use information from other rows j' when estimating diversion in row j.

Consider a hypothetical example in which diversion among three vehicles (Honda Civic, Toyota Prius, and Tesla) is given by a matrix of diversion ratios.

from/to :	*Civic*	*Prius*	*Tesla*
Civic :	50	40	10
Prius :	50	30	20
Tesla :	0	80	20

If we are interested in analyzing a merger between Toyota and Honda, would an estimate of diversion be robust to restricting our attention to only the diversion ratio from Prius to Civic? Or, do we need to estimate diversion from the Prius to all possible alternatives (the entire row)? Do we also need to consider diversion in the other direction (e.g., from Civic to Prius)? Is it empirically important to measure other elements or functions of the matrix (e.g., diversion from Prius to Tesla, or aggregate diversion for the Prius)?

[11] Katz and Shapiro (2003) show that aggregate diversion may be helpful for defining the relevant market.

From a theoretical context, the information we require depends on the calculation we plan to perform. In order to calculate Upward Pricing Pressure (UPP) for a Toyota-Honda merger using equation (1), we only require the diversion ratio from Prius to Civic.[12] For a *partial merger simulation*, as in Hausman, Leonard, and Zona (1994), we calculate the effects of the merger on p_j holding fixed p_{-j} by solving equation (1), and we require the entire row j. In order to perform *full merger simulation* as in Nevo (2001), for which we solve for the system of post-merger equilibrium prices $\mathbf{p}$, we require the entire matrix $D(\mathbf{p})$ in order to calculate the competitive responses for other products in the market.

From an empirical perspective, we may be able to improve our estimates of D_{jk} by using information from the rest of the row ($D_{jk'}$), or information on the sum of the elements of the row ($\sum_{k=0,k\neq j}^{J} D_{jk}$), even though, from a theoretical perspective, the additional information might not be required.

2.2 Diversion as a Treatment Effect

Once we determine the elements or functions of the diversion matrix in which we are interested, we face the related task of choosing how to measure each of the relevant elements, $D_{jk}(p_j, p^0_{-j})$. We consider a hypothetical experiment that raises the price of product j by Δp_j, so that $p_j = p^0_j + \Delta p_j$. We can interpret diversion as a Wald estimator of a treatment effect with a binary treatment (i.e., not purchasing product j) and a binary outcome (i.e., purchasing product k or not). We denote this as:

$$D_{jk}(p_j, p^0_{-j}) = \left|\frac{\Delta q_k}{\Delta q_j}\right| = \left|\frac{q_k(p^0_j + \Delta p_j, p^0_{-j}) - q_k(p^0_j, p^0_{-j})}{q_j(p^0_j + \Delta p_j, p^0_{-j}) - q_j(p^0_j, p^0_{-j})}\right| = \frac{\int_{p^0_j}^{p^0_j+\Delta p_j} \frac{\partial q_k(p_j, p^0_{-j})}{\partial p_j}\, dp_j}{\int_{p^0_j}^{p^0_j+\Delta p_j} \frac{\partial q_j(p_j, p^0_{-j})}{\partial p_j}\, dp_j} \tag{2}$$

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 lower an individual's reservation price for j, the more likely an individual is to receive the treatment. Thus, Δp_j functions as an 'instrument' because it monotonically increases the probability of treatment.

By focusing on the numerator in equation (2), we can re-write the diversion ratio using the marginal treatment effects (MTE) framework of Heckman and Vytlacil (2005), in which

[12] These estimates can be improved by adding the pass-through matrix: $\frac{d\mathbf{p}}{d\mathbf{c}}$. See Jaffe and Weyl (2013) or Miller, Remer, Ryan, and Sheu (2016) for more details.

$D_{jk}(p_j, p_{-j}^0)$ is a marginal treatment effect that depends on p_j.[13]

$$\widehat{D_{jk}^{LATE}}(\Delta p_j) = \frac{1}{\Delta q_j} \int_{p_j^0}^{p_j^0+\Delta p_j} \underbrace{\frac{\partial q_k(p_j, p_{-j}^0)}{\partial q_j}}_{\equiv D_{jk}(p_j, p_{-j}^0)} \frac{\partial q_j(p_j, p_{-j}^0)}{\partial p_j} \partial p_j \tag{3}$$

$$\widehat{D_{jk}^{ATE}} = \frac{1}{\Delta q_j} \int_{p_j^0}^{\overline{p}_j} D_{jk}(p_j, p_{-j}^0) \frac{\partial q_j(p_j, p_{-j}^0)}{\partial p_j} \partial p_j = \left| \frac{q_k(\overline{p}_j, p_{-j}^0) - q_k(p_j^0, p_{-j}^0)}{q_j(\overline{p}_j, p_{-j}^0) - q_j(p_j^0, p_{-j}^0)} \right| \tag{4}$$

As we vary p_j, we measure the weighted average of diversion ratios where the weights $w(p_j) = \frac{1}{\Delta q_j} \frac{\partial q_j(p_j, p_{-j}^0)}{\partial p_j}$ 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. The LATE estimator varies with the size of the price increase because the set of treated individuals varies. Equation (3) confirms that the LATE estimate concentrates more weight near $\mathbf{p^0}$ when demand is more elastic, or when demand becomes increasingly inelastic for larger Δp_j. In equation (4) the average treatment effect (ATE) is the expression for the LATE in which all individuals are treated. This corresponds to an increase in p_j all the way to the choke price $\overline{p}_j$. Evaluating $D_{jk}(p_j^0, p_{-j}^0)$ at pre-merger prices is consistent with a MTE for which Δp_j is infinitesimally small.[14] As we choose larger values for Δp_j our LATE estimate may differ from the MTE evaluated at $D_{jk}(\mathbf{p^0})$.

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}):

$$q_k(p_j + \Delta p_j, p_{-j}) \approx q_k(p_j, p_{-j}) + \frac{\partial q_k(p_j, p_{-j})}{\partial p_j} \Delta p_j + \frac{\partial^2 q_k(p_j, p_{-j})}{\partial p_j^2} (\Delta p_j)^2 + O((\Delta p_j)^3)$$

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

This allows us to compute an expression for the difference between a LATE estimate $\widehat{D_{jk}^{LATE}}(\Delta p_j)$

[13] The MTE is a non-parametric object that 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.

[14] Antitrust authorities also sometimes focus 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%.

and the MTE evaluated at $D_{jk}(\mathbf{p^0})$. We refer to this as the bias of the LATE estimate:

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

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

The expression in equation (6) shows that the bias depends on two things: the magnitude of the price increase Δp_j, and 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 considering small price changes. The disadvantage of a small price change Δp_j is that it also implies that the size of the treated group Δq_j is small. This may increase the variance with which we measure diversion, as shown in equation 7. This is the usual bias-variance tradeoff: a small change in p_j induces a small change in q_j and reduces the bias, but increases the variance; a larger Δp_j (and by construction Δq_j) may yield a less noisy LATE, but may differ from quantity of interest if the antitrust authority is concerned about the potential for small price changes.[15]

Equation (6) also provides insight into the economic implications of assuming a constant treatment effect, such that $D_{jk}(p_j, p_{-j}) = D_{jk}$. Constant diversion requires that the bias calculation in equation (6) is equal to zero. Two functional forms for demand exhibit constant diversion and are always unbiased: linear demand, for which $\frac{\partial^2 q_k}{\partial p_j^2} = 0, \forall j, k$; and 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 Appendix A.1, and show that random-coefficients logit demand, and constant elasticity demands (including log-linear demand) do not generally exhibit constant diversion.

To summarize, we can expect a LATE or ATE measure of diversion to be similar to the MTE evaluated at $\mathbf{p^0}$ when the bias in (6) is small. This happens when: (a) the curvature of demand is low ($\frac{\partial^2 q_k}{\partial p_j^2} \approx 0$), (b) the true diversion ratio is constant (or nearly

[15] Often, the policy question drives the choice of how to empirically measure diversion: is the antitrust authority more concerned about widespread but small price changes, or the potential for product discontinuations? However, equations (6) and (7) illustrate another perspective on the choice of how to empirically measure diversion. Specifically, the elasticity (and super-elasticity) of demand for j may be informative for determining the empirical properties of the diversion measure. If the curvature of demand is steep, a small price increase best avoids bias when measuring diversion at pre-merger prices. However, if sales are highly variable, one may need to consider a larger price increase to reduce variance.

constant) so that $D_{jk}(p_j, p_{-j}) = D_{jk}$, or (c) demand for j is steepest near the market price $\left|\frac{\partial q_j(p_j,p_{-j})}{\partial p_j}\right| \gg \left|\frac{\partial q_j(p_j+\Delta p_j,p_{-j})}{\partial p_j}\right|$.

To illustrate these concepts, figure 1 considers three hypothetical demand curves for Toyota Prius. The first example illustrates diversion to three alternatives (a Honda Civic, Tesla, and an outside good) when demand for a Prius is linear. As price increases from \$25,000 to \$50,000 (along the horizontal axis), diversion to the three alternatives is constant: 63% of potential Prius buyers switch to a Honda Civic, 12% to Tesla, and 25% to the outside option. The histogram along the bottom axis shows the rate at which Prius buyers leave the Prius, which is the rate at which consumers are 'treated' by a price increase. The second example in figure 1 considers diversion for inelastic demand with constant elasticity ($\epsilon = -1$). The rate at which Prius buyers leave is now higher near the market price than at higher prices, so the histogram along the horizontal axis assigns more weight near the market price. Furthermore, the diversion pattern differs as we consider higher price points. There is more substitution to the Honda Civic after a small price increase (90% diversion), and more substitution to the Tesla after a large price increase. Using the histogram to weight diversion across the entire price spectrum provides an ATE estimate of diversion that is 59% to the Honda Civic, 18% to the Tesla, and 22% to the outside good. The third example in figure 1 replicates the second example with an elasticity of $\epsilon = -4$. This greater elasticity changes the relative weighting across different hypothetical price increases, so that more consumers leave at smaller price changes. Although diversion to the three alternatives at any given price point is the same as the case of inelastic demand, the ATE measure of diversion is now more heavily weighted towards consumers that leave at small price changes (72% to Honda Civic, 10% to Tesla, and 18% to the outside good).

2.3 Utilizing Second-Choice Data to Measure Diversion

Often researchers have access to second-choice data. For example, Berry, Levinsohn, and Pakes (2004) observe 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. One may also construct second-choice data through a revealed-preference mechanism by experimentally removing product j from a consumer's choice set for a period of time.[16] One can view such an

[16] Another way to recover second-choice data is to use observational data on consumer choice sets. However, a problem with using observational choice-set variation is that the variation is typically non-random. If one simply compares retail locations that stock product j to locations 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

exogenous product removal as being equivalent to an increase in price to the choke price $\overline{p}_j$, where $q_j(\overline{p}_j, p_{-j}) = 0$. Thus, an exogenous product removal measures the ATE, treating all of the pre-merger consumers of good j and minimizing the variance expression in (7).

Notice the relationship between the ATE measure of diversion $\widehat{D_{jk}^{ATE}}$, and second choice data, where A is the set of available products and $A \setminus j$ denotes the set of available products after the removal of product j:

$$\widehat{D_{jk}^{ATE}} = \left| \frac{q_k(\overline{p}_j, p_{-j}^0) - q_k(p_j^0, p_{-j}^0)}{q_j(\overline{p}_j, p_{-j}^0) - q_j(p_j^0, p_{-j}^0)} \right| = \frac{q_k(\mathbf{p^0}, A \setminus j) - q_k(\mathbf{p^0}, A)}{q_j(\mathbf{p^0}, A)} \tag{8}$$

Under the ATE, all individuals in the population are treated. This has the effect of making the choice of instrument irrelevant in the measure of the treatment effect.

3 Application to Nevo (2000)

In our first application, we use the well-known example from Nevo (2000). This application allows us to measure diversion in two ways: first as the ratio of the estimated derivatives of demand evaluated at pre-merger prices (a MTE evaluated at $\mathbf{p_0}$), and second as the response to a simulated removal of a product (an ATE). The discrete-choice nature of the demand system imposes a 'summing-up constraint' (i.e., that aggregate diversion plus diversion to the outside good sums to one).

The data consist of $T = 94$ markets with $J = 24$ products per market and a $I = 20$ point distribution of heterogeneity for each market. The specification allows for product fixed effects d_j, unobserved heterogeneity in the form of a multivariate normally distributed ν_i with variance Σ, and observable demographic heterogeneity in the form of Π interacted with a vector of demographics d_{it}.

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

We estimate parameters following the MPEC approach of Dubé, Fox, and Su (2012).[17] The

work, Conlon and Mortimer (2013a) establish conditions under which a temporary stock-out event provides random variation in the choice set. The main intuition is that after one conditions on inventory and consumer demand, the timing of a stock-out follows a known random distribution; paired with the assumption that consumer arrival patterns do not respond to anticipated stock-out events, this provides (quasi)-random choice set variation.

[17]Technically we employ the continuously updating GMM estimator of Hansen, Heaton, and Yaron (1996) and adapted to the BLP problem by Conlon (2016). For this dataset, CUE and 2-step GMM produce nearly

estimated coefficient on price is distributed as follows:[18]

$$\beta_{it}^{price} \sim N\left(-62.73 + 588.21 \cdot \text{income}_{it} - 30.19 \cdot \text{income}_{it}^2 + 11.06 \cdot \text{I[child]}_{it}, \sigma = 3.31\right)$$

We denote a measure of diversion evaluated for an infinitesimally small price change as a MTE. We refer to a 'second choice' estimate of diversion as an ATE. For comparison, we also evaluate a Logit model, under which diversion is assumed to be constant. These three treatment effects are defined as:

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

For each of the 94 markets and 24 products, we compute the best substitute for each product-market pair, and calculate the diversion ratio to that product. In table 1, we report these patterns. We find that for MTE and ATE, we get roughly the same amount of substitution on average to the best substitute (around 13-15%). As one might expect, the plain *Logit* fails to capture the closeness of competition and instead finds $9 - 10\%$ substitution on average to the best substitute. We find that the ATE identifies the same best substitute as the MTE around 90% of the time, while the Logit (which identifies the same best substitute for all products) is only in agreement with the MTE 58% of the time. We repeat the exercise and calculate substitution to the outside good in the second panel. We find that the MTE has slightly more outside good substitution ($35 - 37\%$) than the ATE diversion measure ($32 - 34\%$), but far less than the Logit, which predicts that around 54% of consumers switch to the outside good.

One can also compare the different measures of diversion. In table 2, we treat the MTE as the baseline value and compare the difference in the calculated diversion (i.e., the difference between $\log \widehat{D_{jk}^{ATE}} - \log \widehat{D_{jk}^{MTE}}$).[19] The first and third panels of table 2 report this calculation for each product's best substitute and the outside good, similar to table 1. The second panel reports differences for all J substitutes for each product. The ATE measure of diversion is on average $2 - 3\%$ higher than the MTE measure of diversion for

identical point estimates.

[18] One motivation for choosing this particular example is that it demonstrates a large degree of heterogeneity in willingness to pay. In Appendix A.2, we repeat this exercise with a restricted version of the demand model at the original published estimates from Nevo (2000). The restriction imposed is that $\pi_{inc^2,price} = 0$.

[19] As in table 1, an observation is a product-market pair. Table 2 reports means and medians across these $J \cdot T$ observations.

each product's best substitute. Across all substitute products, shown in the second panel, the ATE measure is around $6-8\%$ higher than the MTE measure. When compared to the outside good, the ATE measure is around $8-9\%$ lower than the MTE measure. We also report the mean and median absolute deviation. This indicates that we are both over- and understating substitution on a product-by-product basis, because these are larger in magnitude than the median and mean deviations. As one might expect, the *Logit* model substantially understates (by 40% or more) substitution to the best substitute, as well as substitution to other products (by 25% or more), and overstates substitution to the outside good by about 39%.

The ATE measure may either overstate or understate substitution to other products on average compared to the MTE measure. If the marginal consumer tends to become more (less) inelastic as the price increases, then the ATE will overstate (understate) substitution.[20] Reducing an estimator's ability to accommodate heterogeneity in consumer preferences produces MTE and ATE measures that are closer together. We demonstrate this effect with Monte Carlo simulations of commonly-used parametric demand models in Appendix A.3.

4 Empirical Application to Vending

In our second empirical application, we estimate the ATE form of the diversion ratio using experimental second-choice data. We run a field experiment with multiple treatment arms in which we exogenously remove a product from 66 vending machines located in office buildings in Chicago. The product removals allow us to measure subsequent substitution to the remaining products without any parametric restrictions on demand. We begin with a discussion of the snack foods/vending industry, including potential antitrust issues in 4.1. We discuss our experimental design in 4.2, and describe our experimentally generated data in 4.3.

[20] The elasticity of the marginal consumer will depend on the curvature of demand. For a plain vanilla logit model, the logit error term implies that the elasticity of the marginal consumer increases with price. However, this need not hold for other models of demand. For example, a random coefficient logit model has an inflection point when market share exceeds 0.5. At the market level, $s_j < 0.5$ for all j except for the outside good. Empirically, the outside good share may be less than 0.5 in some markets, but greater than 0.5 in others. For any individual, the predicted share for any product j may exceed 0.5 if the draw from the distribution on the random coefficient is sufficiently high. This detail of the random coefficient logit model implies that one cannot necessarily sign the second derivative of demand, $\frac{\partial^2 q_k}{\partial p_j^2}$, and thus cannot determine theoretically whether the ATE will over- or understate diversion relative to the MTE.

4.1 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*.[21] We report HHI's at both the category level and for all vending products in table 3 from the midwest region of the U.S. 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.

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).[22]

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 revenues since the last service visit on an item-level basis, but do not include time-stamps

[21] 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.

[22] 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.]

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.[23] 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.[24]

4.2 Experimental Design

We implemented 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 (2013b) and Conlon and Mortimer (2017). 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 over a shorter period of time across more machines.[25] 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.[26]

For each treatment, we remove a product from all machines at a client site for a period

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

[24] For example, we combine Milky Way Midnight with Milky Way, and Ruffles Original with Ruffles Sour Cream and Cheddar; and various flavors of Pop-Tarts together.

[25] 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.

[26] 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.

of 2.5 to 3 weeks. The four products that we remove are the two best-selling products from either (a) confections seller Mars Incorporated (Snickers and Peanut M&Ms) or (b) cookie seller Kellogg's (Famous Amos Chocolate Chip Cookies and Zoo Animal Crackers). We refer to exogenously-removed products as the *focal products* throughout our analysis.[27] Whenever a product was exogenously removed, poster-card announcements were placed at the front of the empty product column.[28] 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. Although data are recorded at the level of a service visit, it is more convenient to organize observations by week, because different visits occur on different days of the week.[29] The cost of implementing the experiment consisted primarily of drivers' time.[30]

Our experiment differs somewhat from an ideal experiment. Ideally, we would be able to randomize the choice set for each individual consumer. Technologically, of course, that is difficult in both vending and traditional brick and mortar contexts.[31] 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

[27] Not reported here are two treatment arms removing 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 treatments 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 (2013b) and Conlon and Mortimer (2017).

[28] 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.

[29] 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. In order to define weekly observations, 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. At some site-experiment pairs, weeks run Tuesday to Monday, while others run Thursday to Wednesday.

[30] 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 product removal, and reviewing the data as they were collected. Drivers are generally paid a small commission on the sales on their routes, so if sales levels fell dramatically as a result of the product removals, 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.

[31] 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 product removals.

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 contemporaneous 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.

4.3 Description of Experimental Data

We summarize the data generated by our product removals in table 4. 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 2,700-3,500 individuals, of which around 134-274 would have purchased the focal product in an average week. Each treatment lasts 2.5-3 weeks, and between approximately 14,000-19,000 sales are recorded during the treated periods. The treated group consists of the 400-1,200 individuals who would have purchased the focal product had it been available for 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.

In general, we see that the overall sales per machine-week are higher during the treatment period (between 83.3-89.4) than during the control period (82.2).[32] This illustrates a second challenge, which 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%. This can be seen in Figure 2, which plots the overall sales of all machines from one of the

[32] The per-week sales can be a bit misleading because not all machines are measured in every week during the treatment period. This is because the product removals have slightly different start dates at different client site locations. This leads to a somewhat liberal definition of "treatment week" as only one or two machines might be treated in the final week.

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.

We explore this relationship further in table 5, where we report average sales by week during both the treatment and control periods for key substitutes. The third column reports the quantile that sales during the mean treatment week correspond to in the distribution of control weeks. For example, during the Snickers removal, we recorded an average of 472.5 M&M Peanut sales per week. The average weekly sales of M&M Peanuts was 309.8 units during the control weeks and the treatment average was greater than recorded sales of M&M Peanut during any of our control weeks (100th percentile). Likewise, the overall average weekly sales (across all products) were 5,358 during the treated weeks, compared to a control average of 4,892, which corresponds to the 74.4th percentile of the control distribution for total sales.

5 A Nonparametric Estimator for Diversion

Our estimates of diversion face two major challenges: (1) the set of products may vary for non-experimental reasons across machines and time; (2) demand is volatile both at the product level, and at the aggregate level.

Using four simple assumptions motivated by economic theory, we develop an estimator for the average treatment effect version of the diversion ratio which deals with these challenges. The first two assumptions restrict the set of machine-weeks which can act as a control for a particular machine-treated week as in a *matching estimator*. The second two assumptions involve how estimates of $\widehat{\Delta q_j}, \widehat{\Delta q_k}$ are used to construct $\overline{D}_{jk}$ and employ the principle of *Bayesian Shrinkage*. All four assumptions are implications of the economic restriction that consumers make discrete choices among substitutes.[33] Using the following four assumptions we demonstrate how we estimate $\overline{D}_{jk}$ from our experimental data.

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

Assumption 2. ***Substitutes****: Removing product j can never increase the overall level of sales during a period, and cannot decrease sales by more than the sales of j.*

[33] While we estimate the ATE version of D_{jk} in our example, the procedure described in this section could be used to estimate a LATE if the treatment were a 10% price increase instead of a product removal for example.

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

Assumption 4. ***Unit Simplex***: *$D_{jk} \in [0,1]$ and $\sum_{\forall k} D_{jk} = 1$.*

5.1 Matching Assumptions

Consider an estimate of $\widehat{\Delta q_k}$, where $W = 1$ denotes the removal of product j:[34]

$$\widehat{\Delta q_k} = E[q_k|W = 1] - E[q_k|W = 0]$$

We want to adjust our calculation of the expectation to address the volatility of demand.[35] To be explicit about the problem, one can introduce a covariate ξ (demand shock):

$$E[q_k|W = w] = \int q_k(\xi, w) f(\xi|W = w) d\,\xi$$

The treated and control periods have different distributions of covariates (demand shocks) because $f(\xi|W = 1) \neq f(\xi|W = 0)$. The typical solution involves *matching* or *balancing*, where one re-weights observations in the control period using measure $g(\cdot)$ so that $f(\xi|W = 1) = g(\xi|W = 0)$ and then calculates the expectation $E_g[q_k|Z = 0]$ with respect to measure g.[36] For each treated week t, one can construct a set of matched control weeks within a neighborhood $S(\xi_t)$, where $S(\xi_t)$ is the set of control weeks that correspond to treated week t, and ξ_t is an unobserved demand shock. Having chosen $S(\xi_t)$, the change in sales for the chosen control weeks is given as:

$$\Delta q_{k,t}(\xi_t) = q_{k,t}(\xi_t) - \frac{1}{|\#s \in S(\xi_t)|} \sum_{s \in S(\xi_t)} q_{k,s} \quad \text{with} \quad \widehat{\Delta q_k} = \sum_t \Delta q_{k,t}(\xi_t) \tag{9}$$

Our first two assumptions tell us how to choose $S(\xi_t)$. Assumption 1 is straightforward: it controls for unobserved machine-level heterogeneity by restricting potential controls to different (untreated) weeks at the same machine. If ξ_t were observed one could employ conventional matching estimators (such as k-nearest neighbor or local-linear regression (Abadie and Imbens 2006)). However ξ_t is unobserved, so we rely on Assumption 2 (removing a

[34]One advantage of using a product removal experiment is that $E[q_j|W = 1] = 0$ 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.

[35]Recall, in table 5 treated weeks represent the 74th percentile of the aggregate sales distribution

[36]We omit the usual discussion of the conditional independence assumption because we have randomized assignment of W.

product cannot increase total sales, and cannot reduce sales by more than the sales of the product removed) instead.

We implement Assumption 2 as follows. We let Q_t denote the sales of all products during the treated machine-week, and Q_s denote the overall sales of a potential control machine-week. Given a treated machine-week t, we look for the corresponding set of control periods which satisfy Assumptions 1 and further restrict them to satisfy Assumption 2:

$$\{s : Q_s - Q_t \in [0, q_{js}]\} \tag{10}$$

The problem with a direct implementation of (10) is that periods with (unexpectedly) higher sales of the focal product q_{js} are more likely to be included as a control, which would understate the diversion ratio. We propose a slight modification of (10) which is unbiased. We replace q_{js} with $\widehat{q_{js}} = E[q_{js}|Q_s, W = 0]$. An easy way to obtain the expectation is to run an OLS regression of q_{js} on Q_s using data only from untreated machine-weeks satisfying Assumption 1:

$$S_t \equiv \{s : Q_s^0 - Q_t^1 \in [0, \widehat{b_0} + \widehat{b_1} Q_s^0]\} \tag{11}$$

Thus (11) defines the set of control periods S_t which correspond to treatment period t under our assumptions.[37] Plugging into equation (9) will give us estimates of $\widehat{\Delta q_k}, \widehat{\Delta q_j}$.[38]

5.2 Bayesian Shrinkage Assumptions

Our Assumptions 3 and 4 place restrictions on how we calculate the diversion ratio given our estimates of $\widehat{\Delta q_k}, \widehat{\Delta q_j}$. The idea is that there may be better estimates of $\overline{D}_{jk}$ than the simple ratio $\frac{\widehat{\Delta q_k}}{\widehat{\Delta q_j}}$. For example, we might find large but noisy estimates of diversion to a substitute product based on only a few observations and a better estimate might adjust for

[37] The economic implication of Assumption 2 is that if all treatment and control weeks faced an identical set of substitute products, the sum of the diversion ratios from j to all other products would lie between zero and one (for each t): $\sum_{k \neq j} D_{jk,t} \in [0, 100\%]$.

[38] 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. We find that when all four assumptions are used additional matching criteria have no appreciable effect on our estimates.

that uncertainty.[39]

We can see how these assumptions work by writing the diversion ratio as the probability of a binomial with Δq_j trials and Δq_k successes:

$$\Delta q_k | \Delta q_j, D_{jk} \sim Bin(n = \Delta q_j, p = D_{jk}) \tag{12}$$

This is considered a nonparametric estimator as long as we estimate a separate binomial probability D_{jk} for each (j,k).

We implement Assumption 3 by placing a prior on D_{jk} that restricts all of the mass to the unit interval $D_{jk}|\mu_{jk}, m_{jk} \sim Beta(\mu_{jk}, m_{jk})$. Assumption 4 goes further and restricts the *vector* $\mathbf{D_{j\cdot}}$ to the unit simplex, which we implement with the prior $\mathbf{D_{j\cdot}} \sim Dirichlet(\mu_{j0}, \mu_{j1}, \ldots, \mu_{jK}, m_{jk})$. This has the effect of using information about D_{jk} to inform our estimates for $D_{jk'}$.

There are two ways to parametrize the Beta (and Dirichlet) distributions. In the traditional $Beta(\beta_1, \beta_2)$ formulation β_1 denotes the number of prior successes and β_2 denotes the number of prior failures (observed before any experimental observations). Under the alternative formulation $Beta(\mu, m)$: $\mu = \frac{\beta_1}{\beta_1+\beta_2}$ denotes the prior mean and m denotes the number of "pseduo-observations" $m = \beta_1 + \beta_2$. We work with the latter formulation for both the Beta and Dirichlet distributions.[40] This formula makes it easy to express the posterior mean (under Assumption 3) as a *shrinkage estimator* that combines our prior information with our experimental data:

$$\widehat{D_{jk}} = \lambda \cdot \mu_{jk} + (1-\lambda)\frac{\Delta q_k}{\Delta q_j}, \quad \lambda = \frac{m_{jk}}{m_{jk} + \Delta q_j} \tag{13}$$

The weight put on our prior mean is denoted by λ, and directly depends on how many "pseudo-observations" we observe from our prior before observing 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), $\widehat{D}_{jk}$ shrinks towards μ_{jk} (from either direction). Thus, when our product removals provide lots of information about diversion from j to k we rely on the experimental outcomes, but when our experimental variation is less informative, we rely more on our prior information.[41] This has the desirable

[39] A baseball analogy: If one hitter has 150 hits in 500 at bats and another has 2 hits in 5 at bats, we would likely believe the hitter with the .300 average is a better batter even though he has a lower batting average, while we might think that the second hitter was merely lucky.

[40] The Dirichlet is a generalization of the Beta to the unit simplex. The mean parameters $[\mu_0, \mu_1, \ldots, \mu_k]$ form a unit simplex while m denotes the number of pseudo-observations.

[41] We cannot provide a similar closed-form characterization under Assumption 4. Though there is a conjugacy relationship between the Dirichlet and the Multinomial, there is no conjugacy relationship between

property of taking extreme but imprecisely estimated parameters and pushing them towards the prior mean.

Our remaining challenge is how to specify the prior (μ_{jk}, m_{jk}). Ideally, the location of the prior μ would be largely irrelevant while the prior strength m would be as small as possible.[42] An uniform or uninformative prior might be to let $\mu_{jk} = \frac{1}{K+1}$ where K is the number of substitutes. An informative prior centered on the plain IIA logit estimates would let $\mu_{jk} = \frac{s_k}{1-s_j}$ so that (prior) Diversion is proportional to marketshares.[43]

We use the IIA logit prior not because it is the best estimate of the diversion ratio absent experimental data, but rather because assuming diversion proportional to marketshare is commonplace among practitioners in the absence of better data.[44] An advantage of the shrinkage estimator is that it allows us to nest the parametric estimate of diversion currently used in practice and the experimental outcomes, depending on our choice of m. A smaller m implies a *weaker prior* and more weight on the observed data.

6 Results

6.1 Estimates of Diversion

For each of our four product removals, we report our estimates of the diversion ratios in tables 6 and 7. Along with the number of treated machine-weeks for each substitute, we report the estimates of $\widehat{\Delta q_j}, \widehat{\Delta q_k}$ from our matching estimator under Assumptions 1+2. The next four columns report: the "naive" or "raw" diversion ratio $\widehat{\Delta q_j}/\widehat{\Delta q_k}$, the beta-binomial adjusted diversion ratio under Assumption 3 (with a weak $m = K$ and strong $m = 300$ prior), and the "multinomial" version of the diversion ratio under Assumption 4 with the $m = 4.15$ prior. When we incorporate a prior distribution, we center the mean at the IIA

the Dirichlet and the Binomial except under the special case where the same number of treated individuals Δq_j are observed for each substitute k. For additional discussion regarding prior distributions, please consult Appendix A.4.

[42]Indeed, with all four assumptions this is true. We use only a small number of prior observations $m < 4$, and the location of the prior is almost completely irrelevant.

[43]When μ_{jk} 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 (Chetty, Friedman, and Rockoff 2014) and (Kane and Staiger 2008) or hospital quality (Chandra, Finkelstein, Sacarny, and Syverson 2013).

[44]If we had estimates from a random coefficients demand model, we could use those estimates of the diversion ratio instead. However, we find that under Assumption 4 the choice of μ_{jk} becomes irrelevant. We explore robustness to different priors (including uninformative priors) in Appendix A.4.

logit estimates $\mu_{jk} = \frac{s_k}{1-s_j}$.[45] For each experimental treatment, we report the 12 products with the highest "raw" diversion ratio as well as the outside good.

For Twix, in the second row of table 6, $\Delta q_k = 289.6$ 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.[46] 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. When we apply Assumption 3 with a weak prior ($m_=64$ pseudo-observations, one for each potential substitute), we shrink the estimates of the Non-Chocolate Nestle products ($89.5 \rightarrow 12.4$) much more than Twix ($41.2 \rightarrow 37.9$). When we increase the strength of the prior to $m = 300$ pseudo-observations, we observe even more shrinkage towards the prior mean ($89.5 \rightarrow 3.1$) and ($41.2 \rightarrow 29.5$) respectively.

When we include Assumption 4, we utilize an extremely weak prior with $m = 4.15$ pseudo-observations, but we see substantial shrinkage in our estimates from the *adding up* or *simplex* constraint $\sum_k D_{jk} = 1$ and the constraint that $D_{jk'} \geq 0$ for all k'. We no longer balance large positive diversion to some substitutes with large negative diversion to other substitutes, because negative diversion is ruled out *ex-ante*. This leads to smaller diversion estimates for both Non-Chocolate Nestle ($89.5 \rightarrow 0.7$) and Twix ($41.2 \rightarrow 15.9$). Under Assumption 4, Non-Chocolate Nestle is estimated to hardly be a substitute at all, while Twix remains the second-best substitute behind M&M Peanut, which has a similar "raw" diversion measure, but a larger treated group ($\Delta q_j = -954.3$). Imposing the *adding up* constraint of Assumption 4 also shrinks outside good diversion from $47.5 \rightarrow 23.1$.

We summarize the impact of each of our assumptions in sequence in table 8. To eliminate some noise, we focus only on substitutes where $|\Delta q_j| > 50$. For each treatment arm we report the number of substitutes overall, the number with estimated diversion ratios within certain ranges, and the overall diversion (including the outside good) for all substitutes with positive

[45] For the Dirichlet we add an additional small (uniform) $\frac{1.1}{K+1}$ term to the logit probabilities $m = 3.05$ in order to bound some of the very small prior probabilities (with negative diversion) away from zero. Sampling from zero and near-zero probability events is challenging. Note: this is not required for the Beta distribution because Beta-Binomial conjugacy provides a closed form. Because the marketsize is unobserved, we normalize $\mu_0 = 0.25$ for the outside good. Setting $\mu_0 = 0.75$ gives nearly identical results though requires adding more (uniform) pseudo observations to bound the small probabilities away from zero. See Appendix A.4.

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

estimated diversion ratios $\sum_{k:D_{jk}>0} D_{jk}$, and those with negative estimated diversion ratios $\sum_{k:D_{jk}<0} D_{jk}$. Under just Assumption 1, we find nearly half of products exhibit negative diversion ratios. The sum of diversion ratios for products with positive diversion exceeds 300% while for products with negative diversion it exceeds 200%. Adding Assumption 2 (the matching criteria) produces similar estimates. Aggregated diversion measures are still quite large (-166% to +184%). By accounting for the fact that all four treatments produced overall sales that were above average, this reduces the number of products with very large diversion estimates ($> 20\%$), but also increases the number of products with negative diversion. Assumption 3 eliminates negative diversion estimates by requiring $D_{jk} \in [0, 1]$, and reduces the number of products with extremely large diversion estimates by shrinking them towards the prior (logit) mean. Still, because it doesn't pool any information across substitutes D_{jk} and $D_{jk'}$ the aggregate diversion estimates still exceed 150% (and 250% in three out of four product removals). Only after we include Assumption 4 do we obtain aggregate diversion ratios $< 100\%$.[47] There are now very few products (no more than 3 per experiment) with estimated diversion ratios $> 10\%$.

In table 9, we report the posterior distribution of our preferred diversion estimates under Assumption 4 and the very weak prior $m = 4.15$. We find that in most cases the posterior distribution defines a relatively tight 95% *credible* or *posterior* 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.[48] On the other hand, it demonstrates the power of cross substitute restrictions in Assumption 4; even with a diffuse prior and very little experimental data for some substitutes, requiring diversion to sum to one is sufficient to pin down our estimates.

While Assumption 4 appears relatively innocuous (most researchers are likely willing to assume a multinomial discrete choice framework), one should be cautious precisely because it is so powerful in pinning down the diversion ratio estimates. This suggests that the important empirical decision is determining what the appropriate set of products $\mathcal{K}$ is, such that $\sum_{k\in\mathcal{K}} D_{jk} = 1$. If, for example, one were interested in a merger in which product j acquired both (k, k') but (k, k') were always rotated for one another and never available at the same time, one might want to vary the set of products over which we sum $D_{jk'}$ for each alternative: $\mathcal{K}_k$.[49]

[47] Table 8 does not report exactly 100% because we drop a small number of products with $|\Delta q_j| < 50$ from the summary information in the table.

[48] We compare results with different priors under Assumption 4 in Appendix A.4.

[49] Conlon and Mortimer (2013a) show that assuming all products are always available introduces bias in

One of the perceived benefits of using diversion ratios or UPP alone rather than full merger simulation is that it requires data only from the merging parties, and not from firms outside the merger.[50] The power of Assumption 4 indicates that measuring diversion to all substitute goods (rather than just k) can substantially improve our estimates of D_{jk}.[51] 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.

6.2 Merger Evaluation

An important remedy available to antitrust agencies is that mergers can be approved conditional on the parties divesting a key product or set of products.[52] One can measure whether divesting a product during a merger reduces $\sum_{k \in \mathcal{F}_k} D_{jk}$ below a threshold. Combining equation (1) with assumptions on wholesale prices and marginal costs allows one to interpret diversion ratios in terms of *compensating marginal cost reductions*. Our previous work Conlon and Mortimer (2017) indicates that $p = 0.45$ and $c = 0.15$ are reasonable values of price and cost for confections products. This would imply that an Anti-trust authority would seek a compensating marginal cost reduction of roughly twice the diversion ratio. For example, if an Anti-trust authority expected cost savings from the merger to be 10%, it may request divestitures so that overall diversion to the newly acquired brands was less than 5%.

Our product removals inform us directly about the top two products of Mars (Snickers and Peanut M&Ms) and the top two products of Kellogg's (Zoo Animal Crackers and Famous Amos Chocolate Chip Cookies). This allows us to examine a potential merger between Mars and Kellogg's in both directions. We can measure diversion away from Mars products to Kellogg's brands, and from Kellogg's products to Mars' brands. Tables 10 and 11 analyze diversion and the potential for divested products to restore effective competition. A merger between these two firms is interesting from a policy perspective. Recall that table 3 provides

structural parametric estimates of demand.

[50]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.

[51]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).

[52]Two recent examples are the divestiture of gate slots at specific airports in the American/USAirways merger, and divestiture of the entire U.S. Modelo business during the acquisition of its global activities by Anheuser-Busch InBev.

information on the degree of concentration within the three main product categories of the vending industry (salty snacks, cookies, and confections). If one defines the relevant market at the category level (i.e., confections), then a Mars-Kellogg's merger has no impact on HHI. However, if one defines the market to include all vending products, then a Mars-Kellogg's merger induces a 590 point increase in HHI from a base of 2401.

Table 10 analyzes a potential Mars-Kellogg's merger from the perspective of the Mars products. We report all three measures of diversion from each of the two flagship Mars products to all of Kellogg's brands. The 'adding up' constraint imposed by Assumption 4 substantially reduces overall diversion, with the estimate of total diversion from Snickers to all Kellogg's products decreasing from 50% to 9%, and from 29% to 6% in the case of Peanut M&Ms. Analyzing diversion from Snickers and Peanut M&Ms to Kellogg's product line without Assumptions 3 or 4 indicates that Rice Krispies Treats may be a potential divestiture target, with 27% and 10% diversion respectively. Assumption 4 reduces these estimates to more reasonable levels (1.3% and 2.7% respectively). Products outside of the merging firms with high diversion are not reported in table 10, but are listed in table 6. These include Nestle's Butterfinger, Kraft's Planters Peanuts, and PepsiCo's Rold Gold and Sun Chips snack brands.[53] Recalling the estimates of diversion from table 6, the most important substitutes for Peanut M&Ms are already owned by Mars, as Snickers, Plain M&M's, and Twix comprise diversion of roughly 30%.

Table 11 reports diversion from Kellogg's top two products to Mars' brands. Once again, the 'adding up' constraint of Assumption 4 reduces overall diversion by a factor of three in the case of Zoo Animal Crackers, and corrects a negative estimate of overall diversion in the case of Famous Amos cookies. Estimates of diversion without Assumption 4 identify Milky Way as having high diversion for both of Kellogg's products (23% and 19% diversion under no prior for Animal Crackers and Famous Amos, respectively); applying Assumption 4 reduces these estimates to less than 2%. The degree of diversion from Zoo Animal Crackers to Snickers, Plain and Peanut M&Ms, and Twix Caramel is considerable even with Assumption 4 (a total of about 22%), so one might worry about the potential for a price increase on Zoo Animal Crackers. Products outside of the merging firms that have high diversion from Zoo Animal Crackers and Famous Amos cookies are similar to those for Mars' flagship products: PepsiCo's Rold Gold pretzels (for Animal Crackers), and PepsiCo's Sun Chips and Kraft's Planters peanuts (for Famous Amos).

[53] The estimate of diversion that includes assumption 4 predicts roughly 18% diversion from Snickers to these four products.

7 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. The key input to this approach 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 treatment effect of an experiment in which the price of one product is increased by some 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 field 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 explore the empirical properties of diversion ratios in two applications. In the first, we estimate the discrete-choice demand model from Nevo (2000). In the second, we analyze a randomized field experiment, in which we exogenously remove products from consumers' choice sets and measure the ATE directly.

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. A non-parametric 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, and we find benefits from measuring diversion not only between products involved in a proposed merger, but also from merging products to non-merging products.

Our hope is that this makes a well-developed set of quasi-experimental and treatment effects tools available and better understood to both researchers in industrial organization and antitrust practitioners. While the diversion ratio can be estimated in different ways, 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. (2016): "Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply," *Econometrica*, 84(5), 1961–1980.

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

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. (2016): "The MPEC Approach to Empirical Likelihood Estimation of Demand," Unpublished Mansucript. Columbia University.

Conlon, C., and J. H. Mortimer (2013a): "Demand Estimation Under Incomplete Product Availability," *American Economic Journal: Microeconomics*, 5(4), 1–30.

——— (2013b): "Effects of Product Availability: Experimental Evidence," NBER Working Paper 16506.

——— (2017): "All Units Discount: Experimental Evidence from the Vending Industry," Working Paper.

Das, S. (2017): "Effect of Merger on Market Price and Product Quality: American and US Airways," Working Paper.

Dubé, J.-P. H., J. T. Fox, and C.-L. Su (2012): "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, 80(5), 2231–2267.

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.

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-sample properties of some alternative GMM estimators," *Journal of Business and Economic Statistics*, 14(3), 262–280.

Hausman, J., G. Leonard, and J. D. Zona (1994): "Competitive Analysis with Differenciated Products," *Annales d'Économie et de Statistique*, 34, 159–180.

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.

Josephs, L. (2018): "'We wanted to go first.' Here's what's different in the decade since Delta's merger with Northwest upended the airline industry," *CNBC*.

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.

Katz, M. L., and C. Shapiro (2003): "Critical Loss: Let's Tell the Whole Story," *Antitrust Magazine*.

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., M. Remer, C. Ryan, and G. Sheu (2016): "Pass-Through and the Prediction of Merger Price Effects," *Journal of Industrial Economics*, 64(4), 684–709.

Miller, N., and M. Weinberg (2017): "Understanding the Price Effects of the MillerCoors Joint Venture," *Econometrica*, 85(6), 1763–1791.

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

Nevo, A. (2000): "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand (including Appendix)," *Journal of Economics and Management Strategy*, 9(4), 513–548.

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

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

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

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

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.

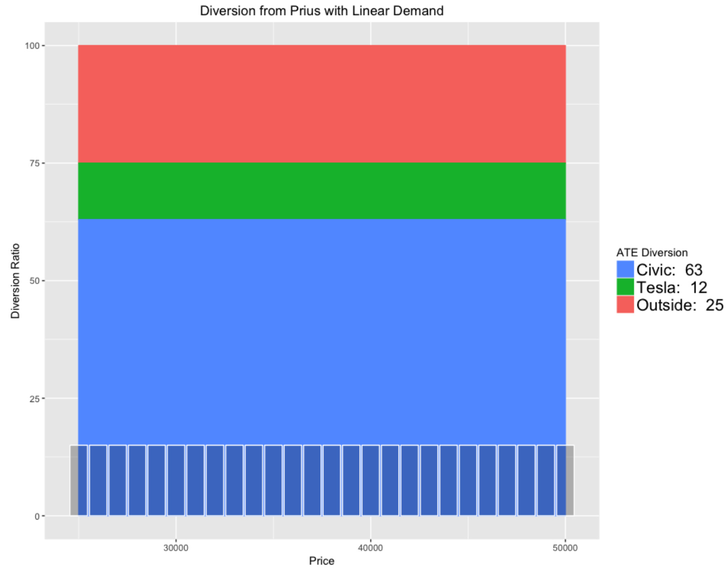
Stock, J. (2010): "The Other Transformation in Econometric Practice: Robust Tools for Inference," *The Journal of Economic Perspectives*, 24(2), 83–94.

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

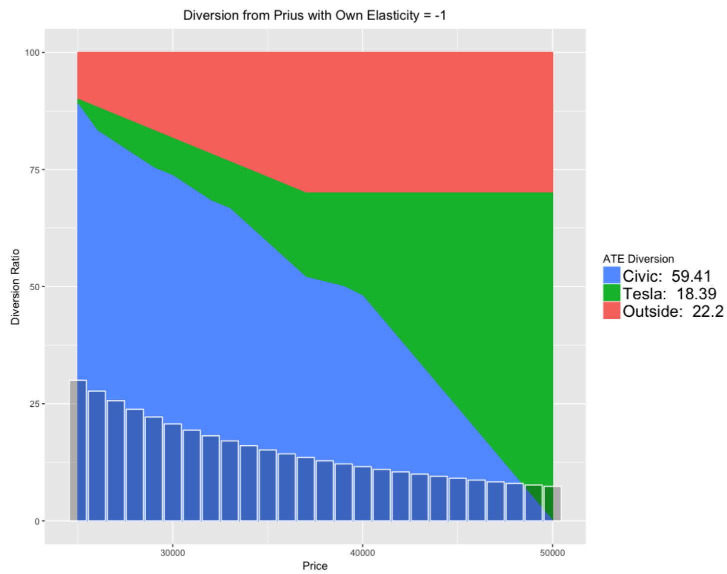
Werden, G. J. (1996): "A Robust Test for Consumer Welfare Enhancing Mergers among Sellers of Differentiated Products," *Journal of Industrial Economics*, 44(4), 409–13.

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

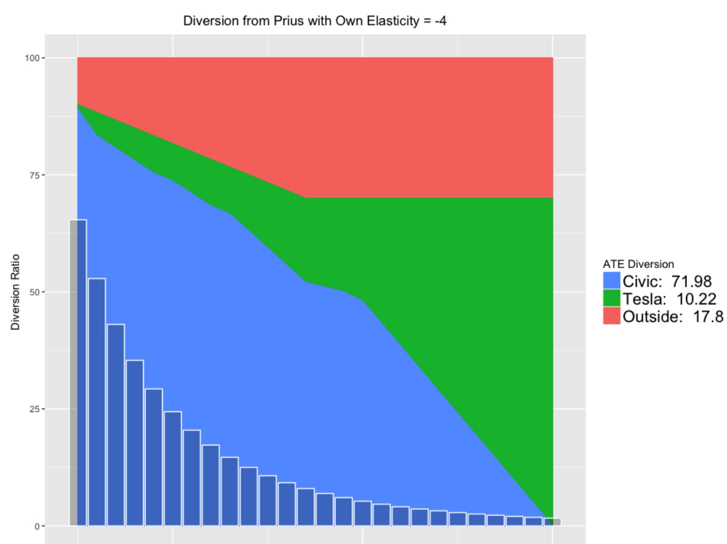
Willig, R. (2011): "Unilateral Competitive Effects of Mergers: Upward Pricing Pressure, Product Quality, and Other Extensions," *Review of Industrial Organization*, 39(1-2), 19–38.



(a) Linear Demand



(b) Inelastic CES Demand



(c) Elastic CES Demand

Figure 1: A Thought Experiment – Hypothetical Demand Curves for Toyota Prius

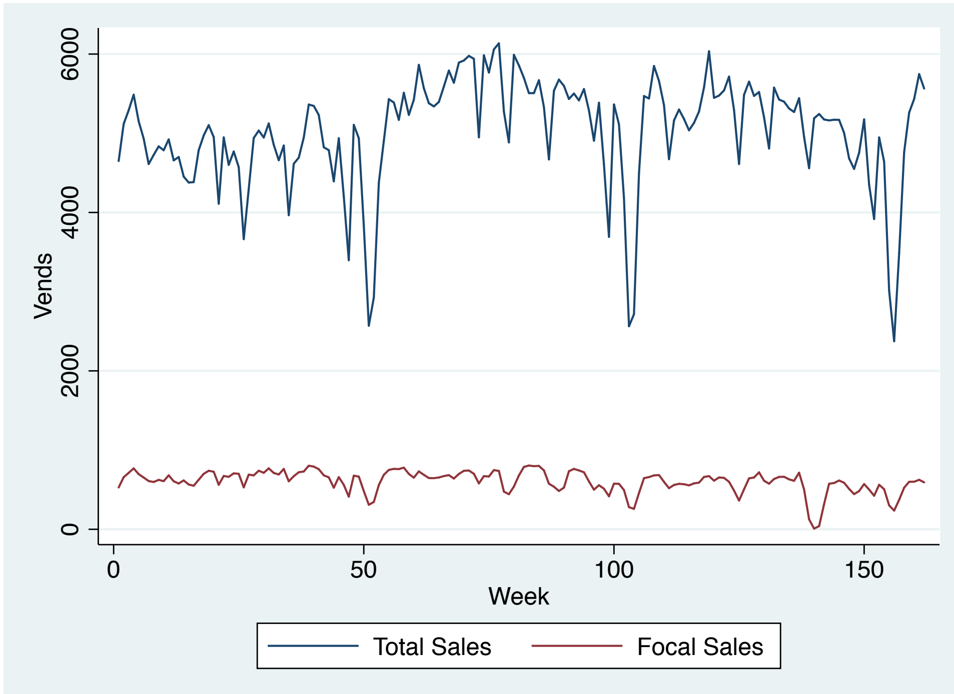


Figure 2: Total Overall Sales and Sales of Snickers and M&M Peanuts by Week

	MTE	*ATE*	*Logit*
	Best Substitute		
Med(D_{jk})	13.26	13.54	9.05
Mean(D_{jk})	15.11	15.62	10.04
% Agree with MTE	100.00	89.98	58.38
	Outside Good		
Med(D_{j0})	35.30	32.40	54.43
Mean(D_{j0})	36.90	33.78	53.46

Table 1: Substitution to Best Substitute and Outside Good

Notes: An observation is a product-market pair. There are 94 markets and 24 products. The first panel reports diversion to each product-market pair's best substitute. The second panel reports diversion to the outside good.

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

Table 2: Relative % Difference in Diversion Measures: Comparison $x = \log(\widehat{D^{MTE}(\mathbf{p_0})})$

Notes: An observation is a product-market pair. There are 94 markets and 24 products. The first panel compares three alternative measures of diversion to the MTE measure for each product-market pair's best substitute. The second panel averages across all possible substitutes. The third panel provides comparisons of the three measures of diversion to the MTE diversion to the outside good.

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 3: 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	M&M Peanut	Famous Amos	Zoo Animal Crackers
# Machines	66	56	62	62	62
# Weeks	160	6	6	5	4
# Machine-Weeks	8,525	223	190	161	167
# Products	76	66	67	65	67
Total Sales	700,404.0	19,005.2	16,232.5	14,394.0	13,910.5
—Per Week	4,377.5	3,167.5	2,705.4	2,878.8	3,477.6
—Per Mach-Week	82.2	85.2	85.4	89.4	83.3
Total Focal Sales*		44,026.3	42,047.8	26,113.2	21,578.4
—Per Week		273.5	262.8	163.2	134
—Per Mach-Week		5.2	4.9	3.1	2.5

Table 4: 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	248.6	260.7	61.6
Snyders	Snyders (Con)	210.2	241.6	52.8
Kellogg	Zoo Animal Cracker	183.1	233.7	96.8
Kraft	Planters (Con)	161.1	218.8	96.0
	Total	4892.1	5357.9	74.4
	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	248.6	251.0	47.2
Mars	Twix Caramel	167.9	213.8	90.4
Kellogg	Zoo Animal Cracker	182.6	198.0	65.6
Pepsi	Baked Chips (Con)	169.4	194.7	68.0
	Total	4886.1	5315.5	65.6
	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	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
	Chocolate Chip Famous Amos 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	248.7	260.7	64.8
Snyders	Snyders (Con)	210.1	236.4	52.8
Pepsi	Sun Chip	150.2	225.5	100.0
Pepsi	Ruffles (Con)	206.9	218.3	62.4
	Total	4890.2	5262.4	64.0

Table 5: 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.

Mfg	Product	Treated Machine Weeks	Δq_k Subst Sales	Δq_j Focal Sales	$\Delta q_k/ \lvert\Delta q_j\rvert$ Div	Assn 3 Diversion ($m = K$)	Assn 3 Diversion ($m = 300$)	Assn 4 Diversion ($m = 4.15$)
				Snickers Removal				
Mars	M&M Peanut	176	375.5	-954.3	39.4	37.0	30.8	18.4
Mars	Twix Caramel	134	289.6	-702.4	41.2	37.9	29.5	15.9
Pepsi	Rold Gold (Con)	174	161.4	-900.1	17.9	16.8	13.9	7.5
Nestle	Butterfinger	61	72.9	-362.8	20.1	17.1	11.2	4.5
Mars	M&M Milk Chocolate	97	71.8	-457.4	15.7	13.8	9.8	4.1
Kraft	Planters (Con)	136	78.0	-759.9	10.3	9.6	7.8	3.8
Kellogg	Zoo Animal Cracker	177	65.7	-970.2	6.8	6.5	5.7	2.9
Pepsi	Sun Chip	159	45.3	-866.1	5.2	5.0	4.3	2.1
Hershey	Choc Hershey (Con)	41	29.8	-179.6	16.6	12.2	6.3	2.0
Kellogg	Rice Krispies Treats	17	17.7	-66.5	26.7	13.5	5.0	1.3
Misc	Farleys (Con)	18	14.9	-114.2	13.0	8.3	3.7	1.0
Nestle	Nonchoc Nestle (Con)	3	9.4	-10.5	89.5	12.4	3.1	0.7
Mars	Choc Mars (Con)	11	6.4	-32.7	19.7	6.5	2.0	0.4
Hershey	Payday	2	1.1	-9.8	10.9	1.4	0.4	0.1
Mars	3-Musketeers	2	0.0	0.0				
Misc	BroKan (Con)	3	0.0	0.0				
	Outside Good	180	460.9	-970.2	47.5			23.1
				M&M Peanut Removal				
Mars	Snickers	218	296.6	-1239.3	23.9	22.9	19.9	16.5
Mars	Twix Caramel	176	110.9	-1014.3	10.9	10.4	8.9	6.8
Mars	M&M Milk Chocolate	99	73.5	-529.6	13.9	12.5	9.2	6.3
Nestle	Raisinets	181	71.8	-1001.1	7.2	6.8	5.8	4.4
Kraft	Planters (Con)	190	61.4	-1046.1	5.9	5.6	4.9	3.6
Hershey	Twizzlers	62	33.0	-333.0	9.9	8.3	5.3	3.4
Kellogg	Rice Krispies Treats	46	22.4	-220.2	10.2	7.9	4.4	2.5
Pepsi	Frito	160	37.2	-902.4	4.1	4.0	3.5	2.4
Misc	Hostess Pastry	11	12.5	-38.6	32.5	12.3	4.0	1.8
Kellogg	Brown Sug Pop-Tarts	10	10.0	-43.5	22.9	9.2	2.9	1.4
Nestle	Nonchoc Nestle (Con)	1	0.9	-4.6	19.5	1.3	0.3	0.2
Misc	Cliff (Con)	1	0.4	-1.8	22.2	0.6	0.1	0.0
	Outside Good	218	606.2	-1238.5	48.9			36.3

Table 6: Raw and Bayesian Diversion Ratios, Mars' Products

Notes: Treated Machine Weeks shows the number of treated machine-weeks for which there was at least one control machine-week. Δq_k Subst Sales shows the change in substitute product sales from the control to the treatment period, while Δq_j Focal Sales shows the analogous change for focal product sales. $\Delta q_j/|\Delta q_j|$ 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, 2 (Substitutes), and 3 (Unit Interval). The weak prior uses the number of products in the choice set during the treatment period, which varies from 64 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, 2 (Substitutes), and Assumption 4 (Unit Simplex).

Mfg	Product	Treated Machine Weeks	Δq_k Subst Sales	Δq_j Focal Sales	$\Delta q_k/$ $\lvert\Delta q_j\rvert$ Div	Assn 3 Diversion ($m = K$)	Assn 3 Diversion ($m = 300$)	Assn 4 Diversion ($m = 4.15$)
	Zoo Animal Crackers Removal							
Pepsi	Rold Gold (Con)	132	114.4	-440.8	25.9	22.9	16.2	9.9
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
Kellogg	CC Famous Amos	144	66.2	-478.2	13.8	12.4	9.1	5.4
Pepsi	Baked Chips (Con)	134	62.5	-447.6	14.0	12.5	9.1	5.3
Mars	Twix Caramel	110	50.2	-339.0	14.8	12.7	8.7	4.6
Sherwood	Ruger Wafer (Con)	119	48.2	-368.7	13.1	11.3	7.6	4.3
Hershey	Choc Hershey (Con)	30	33.6	-132.6	25.3	17.1	7.9	3.8
Kellogg	Rice Krispies Treats	13	23.5	-37.8	62.2	23.2	7.2	3.0
Kar's Nuts	Kar Sweet&Salty Mix	95	30.1	-334.5	9.0	7.7	5.3	2.7
Misc	Popcorn (Con)	56	25.7	-226.9	11.3	8.9	5.1	2.6
Kraft	Planters (Con)	114	28.1	-380.2	7.4	6.5	4.8	2.4
Mars	M&M Milk Chocolate	73	22.7	-295.1	7.7	6.5	4.3	2.2
Kraft	Oreo Thin Crisps	13	14.9	-37.8	39.4	14.7	4.5	1.8
Misc	Hostess Pastry	11	14.7	-62.2	23.7	11.8	4.4	1.8
Misc	Salty United (Con)	6	10.4	-18.9	55.1	12.6	3.4	1.3
Kraft	Chse Nips Crisps	13	8.7	-37.8	23.1	8.6	2.6	1.1
Kar's Nuts	KarNuts (Con)	27	9.2	-85.5	10.8	6.3	2.6	1.0
Mars	Milky Way	9	7.0	-30.8	22.6	7.5	2.2	0.9
Hershey	Payday	2	0.4	-0.4	84.7	0.6	0.1	
	Outside Good	145	240.5	-482.9	49.8			22.0
	Chocolate Chip Famous Amos Removal							
Pepsi	Sun Chip	139	143.6	-355.7	40.4	34.4	22.7	15.7
Kraft	Planters (Con)	121	82.1	-332.6	24.7	20.9	13.7	8.8
Hershey	Choc Hershey (Con)	38	48.6	-66.8	72.7	36.9	13.4	7.2
Pepsi	Frito	119	49.9	-313.2	15.9	13.4	8.9	5.3
Misc	Rasbry Knotts	133	46.6	-345.4	13.5	11.4	7.5	4.8
Pepsi	Grandmas Choc Chip	95	40.0	-259.2	15.4	12.5	7.6	4.5
Pepsi	Dorito Buffalo Ranch	72	38.1	-224.2	17.0	13.3	7.5	4.4
Pepsi	Chs PB Frito Cracker	34	26.9	-83.6	32.1	18.2	7.1	3.7
Kellogg	Choc Sandwich FA	57	28.0	-122.0	22.9	15.1	6.8	3.7
Pepsi	Rold Gold (Con)	147	32.6	-392.2	8.3	7.4	5.5	3.2
Kraft	Oreo Thin Crisps	29	20.7	-43.3	47.9	19.2	6.1	3.1
Mars	Combos (Con)	98	23.6	-274.5	8.6	7.0	4.3	2.6
Nestle	Butterfinger	55	15.9	-152.5	10.5	7.4	3.8	2.0
Mars	Milky Way	26	13.9	-71.6	19.5	10.3	3.9	1.9
Hershey	Twizzlers	40	13.2	-99.3	13.3	8.1	3.5	1.7
Misc	Salty United (Con)	18	9.9	-28.7	34.6	10.7	3.1	1.5
Nestle	Choc Nestle (Con)	1	0.8	-0.3	300.0	1.2	0.3	
Hershey	Payday	2	2.6	6.8	38.9			
Kraft	Fig Newton	2	-0.7	5.7	-12.5			
	Outside Good	156	192.9	-399.1	48.3			21.0

Table 7: Raw and Bayesian Diversion Ratios, Kellogg's Products

Notes: Treated Machine Weeks shows the number of treated machine-weeks for which there was at least one control machine-week. Δq_k Subst Sales shows the change in substitute product sales from the control to the treatment period, while Δq_j Focal Sales shows the analogous change for focal product sales. $\Delta q_j/|\Delta q_j|$ 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, 2 (Substitutes), and 3 (Unit Interval). The weak prior uses the number of products in the choice set during the treatment period, which varies from 64 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, 2 (Substitutes), and Assumption 4 (Unit Simplex).

	Total	Assn 1	Assn 2	Assn 3 ($m = K$)	Assn 4 (m = 4.15)
Snickers Removal					
Products with $D_{jk} < 0$	51	24	26	0	0
Products with $0 \leq D_{jk} \leq 10$	51	13	15	43	48
Products with $10 \leq D_{jk} \leq 20$	51	5	5	5	2
Products with $D_{jk} > 20$	51	9	5	3	1
Sum of all positive D_{jk}s	51	402.84	301.95	265.41	98.72
Sum of all negative D_{jk}s	51	-238.90	-239.07	0.00	0.00
M&M Peanut Removal					
Products with $D_{jk} < 0$	52	20	30	0	0
Products with $0 \leq D_{jk} \leq 10$	52	22	17	48	50
Products with $10 \leq D_{jk} \leq 20$	52	6	3	2	1
Products with $D_{jk} > 20$	52	4	2	2	1
Sum of all positive D_{jk}s	52	295.36	168.92	156.28	97.72
Sum of all negative D_{jk}s	52	-191.73	-157.31	0.00	0.00
Zoo Animal Crackers Removal					
Products with $D_{jk} < 0$	49	11	21	0	0
Products with $0 \leq D_{jk} \leq 10$	49	15	15	39	48
Products with $10 \leq D_{jk} \leq 20$	49	11	8	8	0
Products with $D_{jk} > 20$	49	12	5	2	1
Sum of all positive D_{jk}s	49	644.90	331.96	265.31	92.78
Sum of all negative D_{jk}s	49	-394.12	-280.96	0.00	0.00
Chocolate Chip Famous Amos Removal					
Products with $D_{jk} < 0$	48	25	27	0	0
Products with $0 \leq D_{jk} \leq 10$	48	11	8	37	46
Products with $10 \leq D_{jk} \leq 20$	48	4	7	7	1
Products with $D_{jk} > 20$	48	8	6	4	1
Sum of all positive D_{jk}s	48	417.51	384.60	288.99	95.44
Sum of all negative D_{jk}s	48	-444.17	-400.97	0.00	0.00

Table 8: Summary Statistics for Diversion Estimates across Products

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

Manuf	Product	Mean	2.5^{th} Quantile	25^{th} Quantile	50^{th} Quantile	75^{th} Quantile	97.5^{th} Quantile
				Snickers Removal			
Mars	M&M Peanut	18.40	16.79	17.83	18.39	18.95	20.02
Mars	Twix Caramel	15.88	14.28	15.32	15.88	16.45	17.53
Pepsi	Rold Gold (Con)	7.54	6.49	7.15	7.53	7.92	8.69
Nestle	Butterfinger	4.45	3.53	4.10	4.43	4.78	5.48
Kellogg	Rice Krispies Treats	1.30	0.78	1.09	1.28	1.49	1.95
Nestle	Nonchoc Nestle (Con)	0.67	0.31	0.51	0.65	0.81	1.17
Mars	Choc Mars (Con)	0.44	0.16	0.31	0.42	0.55	0.85
	Outside Good	23.12	21.34	22.50	23.11	23.73	24.91
				M&M Peanut Removal			
Mars	Snickers	16.47	14.83	15.89	16.46	17.04	18.15
Mars	Twix Caramel	6.76	5.60	6.34	6.74	7.16	7.99
Mars	M&M Milk Chocolate	6.26	4.96	5.78	6.25	6.73	7.68
Misc	Hostess Pastry	1.85	1.00	1.49	1.80	2.17	2.95
Kellogg	Brown Sug Pop-Tarts	1.41	0.69	1.10	1.37	1.68	2.39
Nestle	Nonchoc Nestle (Con)	0.15	0.00	0.05	0.11	0.21	0.54
Misc	Cliff (Con)	0.00	0.00	0.00	0.00	0.00	0.03
	Outside Good	36.35	34.21	35.61	36.34	37.09	38.47
				Zoo Animal Crackers Removal			
Pepsi	Rold Gold (Con)	9.89	8.24	9.30	9.88	10.46	11.66
Hershey	Choc Hershey (Con)	3.81	2.66	3.35	3.77	4.22	5.17
Kellogg	Rice Krispies Treats	2.99	1.93	2.56	2.95	3.36	4.28
Kraft	Oreo Thin Crisps	1.85	1.04	1.51	1.81	2.14	2.88
Misc	Hostess Pastry	1.80	1.02	1.47	1.76	2.08	2.79
Misc	Salty United (Con)	1.25	0.61	0.97	1.21	1.49	2.12
Kraft	Chse Nips Crisps	1.10	0.51	0.83	1.06	1.32	1.91
	Outside Good	21.98	19.64	21.15	21.96	22.78	24.43
				Chocolate Chip Famous Amos Removal			
Pepsi	Sun Chip	15.75	13.53	14.94	15.72	16.52	18.11
Kraft	Planters (Con)	8.75	7.04	8.13	8.72	9.35	10.64
Hershey	Choc Hershey (Con)	7.18	5.38	6.49	7.14	7.83	9.21
Pepsi	Chs PB Frito Cracker	3.74	2.51	3.25	3.70	4.19	5.21
Kellogg	Choc SandFamous Amos	3.69	2.47	3.21	3.65	4.12	5.15
Kraft	Oreo Thin Crisps	3.05	1.90	2.59	3.01	3.47	4.47
Misc	Salty United (Con)	1.47	0.70	1.13	1.42	1.75	2.49
	Outside Good	20.95	18.43	20.05	20.94	21.83	23.57

Table 9: Posterior Distribution of Dirichlet $\alpha_{jk} = \frac{s_j}{1-s_j} * 3.05 + \frac{1}{K+1} * 1.1$, $m_{jk} = 4.15$

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

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

Table 10: Divestitures: Diversion from Mars to Kellogg's products.

* Combines Strawberry, Cherry, and Brown Sugar flavors.
Notes: Number of observations for outside good reflects total treatment weeks.
95% credible intervals given in parentheses for binomial and multinomial diversions.

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

Table 11: Divestitures: Diversion from Kellogg's to Mars' products.

* Combines Skittles, Starburst and other non-chocolate Mars candies.
Notes: Number of observations for outside good reflects total treatment weeks.
95% credible intervals given in parentheses for binomial and multinomial diversions.

A Appendix:

A.1 Diversion Under Parametric Demands

This section derives explicit formulas for the diversion ratio under common parametric forms for demand, focusing on whether or not the diversion ratio implied by a particular parametric form of demand is constant with respect to the magnitude of the price increase. We show that the IIA logit and linear demands model exhibit this property, while the log-linear and mixed logit models 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. We specify linear demand as:

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

where β_{kj} is the increase or decrease in k's quantity due to a one-unit increase in prouct j's price. This implies a diversion ratio corresponding to a change in price p_j of Δp_j:

$$D_{jk} = \frac{\Delta Q_k}{\Delta Q_j} = \frac{\beta_{kj} \Delta p_j}{\beta_{jj} \Delta p_j} = \frac{\beta_{kj}}{\beta_{jj}} \tag{A.14}$$

Thus, 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 demand, diversion is a global property. Any magnitude of price increase evaluated at any initial set of prices and quantities will result in the same measure of 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:

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

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

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

$$D_{jk} \approx \frac{Q_k(\mathbf{p})}{Q_j(\mathbf{p})} \cdot \frac{\varepsilon_{kj}}{\varepsilon_{jj}} \tag{A.15}$$

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, which implies that the diversion ratio does not depend on the magnitude of the price increase. We consider two price increases: an infinitesimal one and an increase to the choke price $p_j \to \infty$. The derivation of the diversion ratio D_{jk} under an IIA logit demand model uses a utility specification and choice probabilities given by well-known equations, where a_t denotes the set of products available in market t:

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

Under logit demand, an infinitesimal price change in p_j exhibits identical diversion to setting $p_j \to \infty$ (the choke price). For an infinitesimally small price change,

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

For a price change to the choke price,

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

In both cases, diversion is the ratio of the change in the marketshare of the substitute good divided by the share of consumers no longer buying the focal good (under the initial set of prices and product availability). It does not depend on any of the estimated parameters (α, β). The bias expression for the diversion ratio is equal to zero under IIA logit demand (i.e., $D_{jk} = -\frac{\partial^2 q_k}{\partial p_j^2} / \frac{\partial^2 q_j}{\partial p_j^2}$), shown here:

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

Nested Logit Demand
Recall the estimating equation for the nested logit from Berry (1994):

$$\ln s_{jt} - \ln s_{0t} = x_{jt}\beta - \alpha p_{jt} + \sigma \ln s_{j|g,t} + \varepsilon_{jt}$$

The derivatives of marketshare with respect to price are given by:

$$
\begin{aligned}
\frac{\partial s_j}{\partial p_j} &= \alpha s_j \left(\frac{-1}{1-\sigma} + \frac{\sigma}{1-\sigma} s_{j|g} + s_{jt} \right) \\
\frac{\partial s_k}{\partial p_j} &= \left\{ \begin{array}{ll} \alpha s_j \left(\frac{\sigma}{1-\sigma} s_{k|g} + s_{kt} \right) & \text{for } (j,k) \text{ in same nest} \\ \alpha s_j s_k & \text{otherwise} \end{array} \right\}
\end{aligned}
$$

The interesting case is when both products are in the same nest. The diversion ratio is given by:

$$
\begin{aligned}
\widehat{D}_{jk} = -\frac{\frac{\partial s_k}{\partial p_j}}{\frac{\partial s_j}{\partial p_j}} &= -\frac{\frac{\sigma}{1-\sigma} s_{k|g} + s_k}{\frac{-1}{1-\sigma} + \frac{\sigma}{1-\sigma} s_{j|g} + s_j} = -\frac{\frac{\sigma}{1-\sigma} s_{k|g} + s_{k|g} s_g}{\frac{-1}{1-\sigma} + \frac{\sigma}{1-\sigma} s_{j|g} + s_{j|g} s_g} = -\frac{s_{k|g}\left(\frac{\sigma}{1-\sigma} + s_g\right)}{\frac{-1}{1-\sigma} + s_{j|g}\left(\frac{\sigma}{1-\sigma} + s_g\right)} \\
&= \frac{s_{k|g}}{\underbrace{\left(\frac{1}{\sigma + (1-\sigma)s_g}\right)}_{=Z(\sigma, s_g)} - s_{j|g}}
\end{aligned}
$$

As $\sigma \to 1$ everyone stays within the group and $D_{jk} = \frac{s_{k|g}}{1-s_{j|g}}$.
As $\sigma \to 0$ we collapse to the logit and $D_{jk} = \frac{s_{k|g}}{\frac{1}{s_g} - s_{j|g}} \cdot \frac{s_g}{s_g} = \frac{s_k}{1-s_j}$.
For all other values, $D_{jk} = \frac{s_{k|g}}{Z(\sigma, s_g) - s_{j|g}}$ where $Z(\sigma, s_g) \geq 1$.

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 induce 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:

$$u_{ijt} = \overbrace{\underbrace{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}} + \mu_{ijt}}^{V_{ijt}} + \varepsilon_{ijt}$$

Using the chain rule (for an arbitrary z_{jt}) we can write:

$$\frac{\partial V_{ijt}}{\partial z_{jt}} = \frac{\partial V_{ijt}}{\partial \delta_{jt}} \cdot \frac{\partial \delta_{jt}}{\partial z_{jt}} + \frac{\partial V_{ijt}}{\partial \mu_{ijt}} \cdot \frac{\partial \mu_{ijt}}{\partial z_{jt}}$$

Absent taste heterogeneity for z_{jt} we have that $\frac{\partial \mu_{ijt}}{\partial z_{jt}} \equiv 0$ and $\frac{\partial V_{ijt}}{\partial z_{jt}} = 1 \cdot \frac{\partial \delta_{jt}}{\partial z_{jt}} = \beta_z$. When consumers have a common price parameter $\frac{\partial V_{ik}}{\partial p_j} = \alpha$,

$$\widehat{D_{jk}^{LATE}} = \frac{\frac{\partial S_k}{\partial p_j}}{\left|\frac{\partial S_j}{\partial p_j}\right|} = \frac{\int s_{ij} s_{ik} \frac{\partial V_{ik}}{\partial p_j}}{\int s_{ij}(1 - s_{ij}) \frac{\partial V_{ij}}{\partial p_j}} \to \frac{\int s_{ij} s_{ik}}{\int s_{ij}(1 - s_{ij})} \tag{A.16}$$

$$\widehat{D_{jk}^{ATE}} = \frac{\int \frac{e^{V_{ik}}}{1 + \sum_{l \in a \setminus j} e^{V_{il}}} - \frac{e^{V_{ik}}}{1 + \sum_{l' \in a} e^{V_{il'}}}}{\int - \frac{e^{V_{ij}}}{1 + \sum_{l \in a} e^{V_{il}}}} = \frac{1}{s_j} \int \frac{s_{ij} s_{ik}}{(1 - s_{ij})} \tag{A.17}$$

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 the 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 $\frac{\partial q_k}{\partial p_j} / |\frac{\partial q_j}{\partial p_j}|$.

For any model within the logit family, it should be clear that the ATE form of the diversion ratio does not depend on the price "instrument" (A.17), as long as we drive the purchase probability to zero. Second choice data don't depend on whether price is increased or quality (or some component thereof) is decreased. When the entire population (buyers of j) is treated, the instrument that selects individuals into treatment does not matter.

Likewise, because the random coefficients model is a single index model, any z_{jt} which affects only the mean utility component δ_{jt} and not the unobserved heterogeneity μ_{ijt} yields the same marginal diversion $\widehat{D}_{jk}$. This can be seen in (A.16) which does not depend on $\partial V_{ijt} / \partial p_{jt}$. This has the advantage that the (marginal/infinitesimal) diversion ratio can be identified in the random coefficients logit model even when a (common) price parameter

α is not identified. The easiest choice of a non-price z_{jt} is ξ_{jt}, the unobserved product quality term. The role of β_z is to determine how many individuals receive the treatment as we vary the instrument, but this matters neither in the infinitesimal case, nor in the ATE (second-choice) case.

It is important to note that for any two variables for which there is no preference heterogeneity, they yield the same infinitesimal diversion ratios under the logit family. Likewise any two variables (irrespective of preference heterogeneity) yield the same ATE (second choice diversion ratios). This is in contrast with the treatment effects literature, where different instruments trace out different MTEs. Thus, the single index of the logit family places an important restriction on the treatment effects (which may or may not be reasonable).

A.2 Alternative Specifications for Nevo (2000) Example

Here we repeat the same exercise as in section 3 from the text, but with different parameter estimates. In the first case we use the original published estimates from Nevo (2000) where β_{it}^{price} exhibited substantially less heterogeneity, while in the second we consider a restricted MPEC estimator which imposes the demographic interaction between $income^2$ and $price$ is equal to zero: $\pi_{inc^2,price} = 0$. We report those parameter estimates below as well as the estimates in the text from Dubé, Fox, and Su (2012):

$$\begin{aligned}
\text{DFS (2012):}\ \beta_{it}^{price} &\sim N(-62.73 + 588.21 \cdot \text{income}_{it} \quad -30.19 \cdot \text{income}_{it}^2 \quad +11.06 \cdot \text{I[child]}_{it}, \sigma = 3.31)\\
\text{Nevo(2000):}\ \beta_{it}^{price} &\sim N(-32.43 + 16.60 \cdot \text{income}_{it} \quad -0.66 \cdot \text{income}_{it}^2 \quad +11.63 \cdot \text{I[child]}_{it}, \sigma = 1.85)\\
\text{Restricted:}\ \beta_{it}^{price} &\sim N(-34.09 + 8.53 \cdot \text{income}_{it} \quad +18.16 \cdot \text{I[child]}_{it}, \sigma = 1.04)
\end{aligned}$$

We report both cases in table 12. We observe substantially less heterogeneity in β_{it}^{price} and we also observe that the MTE and ATE measures tend to be more similar to one another.

A.3 Discrepancy Between Average and Marginal Treatment Effects

We perform a Monte Carlo study to analyze the extent to which the average treatment effect deviates from the marginal treatment effect under different demand specifications. 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 (2016), Judd and Skrainka (2011) and Conlon (2016), in which prices are endogenously determined via a Bertrand-Nash game given the other utility parameters (rather than directly drawn from a 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 $\eta_j = \rho\omega_{j1} + (1-\rho)\omega_{j3} - 1$ and $(\omega_1, \omega_2, \omega_3) \sim^{i.i.d.} U[0,1]$. Following Armstrong (2016) and Conlon (2016), 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 according to a multi-product Bertrand-Nash equilibrium.

	med$(y-x)$	mean$(y-x)$	med$(\|y-x\|)$	mean$(\|y-x\|)$	std$(\|y-x\|)$
			Nevo (2000) Estimates		
			Best Substitutes		
ATE	1.39	2.45	2.51	4.16	5.00
Logit	-31.83	-35.01	32.72	38.40	29.13
			All Products		
ATE	1.05	1.91	3.18	4.97	5.42
Logit	-29.15	-29.09	33.98	40.05	31.60
			Outside Good		
ATE	-2.90	-3.24	3.09	3.76	3.05
Logit	32.52	40.49	32.52	41.02	30.67
			Restricted Estimates $\pi_{inc^2,price} = 0$		
			Best Substitutes		
ATE	2.52	5.26	4.77	7.78	9.02
Logit	-41.56	-40.50	43.23	47.47	29.60
			All Products		
ATE	2.02	3.11	7.34	11.12	11.39
Logit	-33.48	-19.13	50.80	56.00	36.46
			Outside Good		
ATE	-5.11	-6.45	5.14	6.70	5.73
Logit	30.46	35.38	30.56	37.05	27.04

Table 12: Alternative Specifications for Nevo (2000).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
μ	0.1	0.1	1	1	2	3	3
σ	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 Elasticity	-5.37	-3.50	-4.64	-3.12	-3.70	-4.41	-1.93
'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
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

Table 13: Simulation comparing ATE and MTE for Random Coefficients Logit

Notes: Results reported for 100 trials.

α	-1.000	-1.000	-1.000	-1.000	-4.000	-4.000	-4.000	-4.000
σ_x	0.500	1.000	2.000	3.000	0.500	1.000	2.000	3.000
s_0	0.134	0.149	0.182	0.223	0.989	0.987	0.975	0.951
own elasticity	-1.476	-1.459	-1.431	-1.399	-7.735	-7.718	-7.811	-7.895
Avg max Discrepancy	2.538	2.578	2.398	2.447	0.151	0.236	0.859	1.010
Std max Discrepancy	0.878	0.971	0.836	0.987	0.059	0.087	0.340	0.851
'Worst Case' Avg ATE	13.019	13.328	13.572	16.109	1.185	2.090	6.923	10.958
'Worst Case' Avg MTE	10.490	10.782	11.255	13.920	1.034	1.854	6.064	9.973
% deviation	26.241	25.913	23.570	19.643	15.557	13.608	15.475	12.682

Table 14: Monte Carlo Simulations

Notes: σ_p is set to 0.5 for all of these runs. Additional results are available that vary σ_p from 0 to 1. As α increases, demand becomes more elastic. As σ_x increases, consumer tastes become more heterogeneous.

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. The first two rows of table 13 report the average outside good share and the average own price elasticity for each simulated market as we vary μ from 0.1 to 3 and σ from 0.5 to 2. The next two rows report the ATE and MTE measures for the pair of products in each trial with the largest discrepancy between the two measures. The last two rows report the discrepancy between the two measures of diversion.

Although there are some simulations where the $ATE < MTE$, the vast majority of simulations for the random coefficients model with a lognormally distributed price coefficient implies that using the stock-out based ATE overstates the 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).

Table 14 reports detail from additional Monte Carlo simulations. When price elasticity is high, the discrepancy between the two measures is small. Elastic demand drops off quickly, implying that a small change in price has a similar effect as a product removal. This is most easily seen by comparing the left panel of table 14 with the right panel. Heterogeneous tastes, on the other hand, lead to larger discrepancies between the ATE and MTE measures. This is most easily seen in table 14 by moving across the columns from left to right of each panel.

A.4 Robustness to Alternative Priors Under Assumpton 4

Our formulation of Assumption 4 uses Dirichlet prior centered on the IIA logit diversion estimates (proportional to marketshare).[54] Because some potential substitutes see $\Delta q_k \leq 0$ and may have priors s_k near zero, we need to bound the prior probabilities away from zero in order to avoid drawing from degenerate distributions. Therefore we add 1.1 pseudo observations from a uniform prior $\frac{1}{K+1}$ to each substitute. This gives a Dirichlet parameter of $\alpha_k = \frac{s_k}{1-s_j} + \frac{1.1}{K+1}$. We then choose α_0 so that outside good share $\mu_0 = 0.25$ for the prior distribution. This results in $m = 4.15$ pseduo observations for our (very weak) prior distribution.

For robustness we consider two other (weak) priors. For one, we keep everything else the same but choose $\mu_0 = 0.75$. For the other we consider the "uninformative" or uniform prior of $\alpha_k = \frac{1}{K+1}$ with $m = 1.1$ pseudo observations. An additional approach is to choose a prior distribution that more closely resembles a logit model. This approach is known as the *over-parametrized normal* which is a common technique in the statistics literature and is better behaved for rare events. See Gelman, Bois, and Jiang (1996) and Blei and Lafferty (2007).[55]

Alternative Assumption. *"Unit Simplex": $D_{jk} \in [0,1]$ and $\sum_{\forall k} D_{jk} = 1$*
$\Delta q_k | \Delta q_j, D_{jk} \sim Bin(n = \Delta q_j, p = D_{jk})$ and $\eta_{jk} | \mu_{jk}, \sigma_{jk} \sim N(\mu_{jk}, \sigma_{jk})$, $D_{jk} = \frac{\exp[\eta_{jk}]}{\sum_{k'} \exp[\eta_{jk'}]}$

For each of these specifications we report the maximum absolute deviations for the posterior mean of the estimated diversion ratios $L_\infty = \max_k |\hat{D}_{jk} - \tilde{D}_{jk}|$ where the base $\tilde{D}_{jk}$ is given by our Dirichlet prior centered on the (adjusted) IIA logit estimates $\alpha_k = \frac{s_k}{1-s_j} + \frac{1.1}{K+1}$ with $m = 4.15$ pseudo-observations. The discrepancies between these priors are reported in table 15. We obtain nearly identical results (differences less than 0.03 *percentage points*) when compared to the Dirichlet with a uniform $\frac{1}{K+1}$ prior and $m = 1.1$ pseudo-observations and the multinomial logit transformed normal prior. There is a somewhat larger discrepancy (differences less than 0.5 *percentage points*) when compared to a somewhat stronger ($m = 9.6$) Dirichlet prior with a larger outside good share $\mu_0 = 0.75$, which we attribute to the stronger prior rather than the share of the outside good.[56]

[54] One can transform the Dirichlet as follows: $Dirichlet(\alpha_0, \ldots, \alpha_K)$ has $\mu_k = \frac{\alpha_k}{m}$ and $m = \sum_{k'=0}^{K} \alpha_{k'}$.

[55] We can interpret this as a multinomial logit model with product intercepts η_{jk} which are estimates with some sampling error σ_{jk}. However, as σ increases, because the multinomial logit transformation is nonlinear this tends towards $\mu_{jk} = \frac{1}{K+1}$.

[56] We need a somewhat stronger prior to bound small probabilities away from zero when the outside good share is larger.

Experiment	Dirichlet	Dirichlet	Normal-Logit
Prior Mean $\mu_j =$	$\frac{s_k}{1-s_j}$, $(s_j = 0.75)$	$\frac{1}{K+1}$	$\frac{1}{K+1}$
Prior Strength	$m = 9.60$	$m = 1.1$	$\sigma^2 = 100$
Snickers	0.218	0.023	0.017
Zoo Animal Crackers	0.428	0.020	0.014
Chocolate Chip Famous Amos	0.537	0.033	0.028
M&M Peanut	0.216	0.014	0.025

Table 15: Maximum Absolute Deviation (percentage points) between Dirichlet parametrized by (adjusted) IIA logit shares ($\alpha_k = \frac{s_k}{1-s_j} + \frac{1.1}{K+1}$, $\mu_0 = 0.25$, $m = 4.15$) and alternatives.

A.5 Additional Results

Mfg	Product	Treated Machine Weeks	Avg # Cntl Per Trt	Δq_k Subst Sales	Δq_j Focal Sales	Δq_k/ $\lvert\Delta q_j\rvert$ Div	Treated Machine Weeks	Avg # Cntl Per Trt	Δq_k Subst Sales	Δq_j Focal Sales	Δq_k/ $\lvert\Delta q_j\rvert$ Div
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 Hershey (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	187	120.3	132.0	-923.6	14.3	177	9.5	65.7	-970.2	6.8
Kellogg	Choc Sandwich FA	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	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	113	116.6	50.1	-565.7	8.8	104	8.9	27.6	-597.1	4.6
Kellogg	CC 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	137	93.3	34.9	-666.8	5.2	129	7.2	13.0	-686.5	1.9
Pepsi	Dorito Buffalo Ranch	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	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	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	44	78.2	-1.3	-186.6	-0.7	39	9.6	-7.3	-174.0	-4.2
Kellogg	Strwbry Pop-Tarts	162	118.1	-6.0	-792.7	-0.8	154	9.9	-40.5	-819.4	-4.9
Gen Mills	Nature Valley	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	Chse Nips Crisps	20	93.8	-1.1	-72.9	-1.5	17	6.5	-6.3	-66.5	-9.4
Pepsi	Smartfood	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	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	69	124.8	-15.4	-371.5	-4.1	66	9.1	-22.3	-372.9	-6.0
Pepsi	Grandmas CC	119	114.6	-29.9	-589.7	-5.1	111	9.8	-36.3	-580.7	-6.3
Kraft	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	154	106.0	-69.5	-749.8	-9.3	144	9.4	-84.4	-798.1	-10.6
Gen Mills	Oat n Honey Granola	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	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 16: Simple Matching Estimator (with and without Assumption 2) (Snickers Removal)

Notes: Trt'd Mach-Weeks reports 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. Δq_k shows the change in substitute product sales from the control to the treatment period.

Manuf	Product	Δ Focal Sales	No Prior	Beta-Bin Diversion $m = J^\dagger$	Beta-Bin Diversion $m = 150$	Beta-Bin Diversion $m = 300$	Beta-Bin Diversion $m = 600$
			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 Hershey (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	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 Hershey (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	Chse Nips Crisps	-37.8	23.1	8.6	4.7	2.6	1.4
			Chocolate Chip Famous Amos Removal				
Nestle	Choc Nestle (Con)	-0.2	300.0	1.2	0.6	0.3	0.2
Hershey	Choc Hershey (Con)	-66.8	72.7	36.9	22.5	13.4	7.4
Kraft	Oreo Thin Crisps	-43.3	47.9	19.2	10.8	6.1	3.3
Pepsi	Sun Chip	-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 17: 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 1,2, and 3 (Unit Interval), using different number of pseudo-observations.

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

A.6 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)-(4).

```
% Main Specification: Dirichlet Prior
data {
  int<lower=1> J;               // number of products, including outside good
  int<lower=1> N[J];            // number of trials
  int<lower=0> y[J];            // number of successes for each product j
  vector[J] priors;    // mean of the distribution of alpha
}

parameters {
  simplex[J] theta;
}

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

% Alternative Specification: Multinomial Logit/Normal
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]);
    }
}
```