An Experimental Approach to Merger Evaluation *

Christopher T. Conlon†
Julie Holland Mortimer‡

February 8, 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 elasticities, and second-choice data. Policy-makers may be interested in either, depending on context. For example, small price increases may be important in the retail gasoline or airline industries, whereas second-choice data (or large price changes) may be relevant for understanding product discontinuations or mergers between hospital systems. The former can be interpreted as a Local Average Treatment Effect; the latter as an Average Treatment Effect. Although experimental variation in pricing can be used to recover a diversion function that varies with the size of a price increase, this is rarely feasible. We analyze the impacts of several hypothetical mergers using observational price variation and experimental variation in second-choice data.

*We thank Mark Stein, Bill Hannon, and the drivers at Mark Vend Company for implementing the experiments used in this paper, providing data, and generally educating us about the vending industry. We thank Dennis Carlton, Joe Farrell, Ken Hendricks, Aviv Nevo, Dan O'Brien, Bill Rogerson, Ralph Winters, and participants at the FTC Microeconomics Conference, 'I.O. Fest' at UC Berkeley, and the Antitrust conference at the University of British Columbia for comments. Andrew Copland, Bogdan Genchev, Tom Gole, Adam Kapor, Joonwoo Park, and Sharon Traiberman provided exceptional research assistance. Financial support for this research was generously provided through NSF grant SES-0617896. Any remaining errors are our own.

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

‡Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467, and NBER. email: julie.mortimer.2@bc.edu

1 Introduction

The authority to challenge proposed mergers is a key tool for enforcing antitrust policy. In the U.S., two agencies, the Department of Justice (DOJ) and the Federal Trade Commission (FTC) are charged with this task, and an important consideration for the agencies is how to balance clarity in their merger approval policies – so that firms can pursue potential transactions efficiently – with flexibility – so that the agencies can account for important facts that may be unique to the review of any given case. The agencies have endeavored to strike this balance by publishing merger guidelines, which aim to describe the process by which the agencies conduct reviews of proposed mergers, particularly horizontal mergers.[1]

In their merger guidelines, as well as in practice, the agencies identify the concept of *unilateral* effects as being important to the review of proposed mergers. 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. A concept that is critical to analyzing the potential for unilateral effects is *diversion.* The current 2010 merger guidelines define a diversion ratio as "the fraction of unit sales lost by the first product due to an increase in its price that would be diverted to the second product." The guidelines note that:

> *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, anti-trust 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 greater.

A related concept is *aggregate diversion*, 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." Relatedly, one can define a 'diversion matrix,' as a matrix whose elements report diversion between each pair of products in the candidate

[1]The first set of merger guidelines was published by the DOJ in 1968, and updated in 1982 and 1984. In 1992, the DOJ and the FTC jointly issued new guidelines, which were updated in 1997, and more recently in 2010. For discussion of theoretical and empirical advances in horizontal merger review, see Whinston (2006).

[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.

market; diagonal elements are defined as diversion to the outside good.[3] The 2010 merger guidelines emphasize the importance of understanding the diversion ratio(s) between the products of the merging firms for analyzing a merger. However, as Katz and Shapiro (2003) points out, other market information – in particular, aggregate diversion – may also play an important role in understanding the potential impacts of a merger. Policy makers need to consider (i) which elements of a diversion matrix are important to include in a merger evaluation, and (ii) how each relevant element of the diversion matrix should be measured. The best approach may depend on the context of the particular merger.

When determining the elements of a diversion matrix that are important for analyzing a particular merger, one can consider each individual element, or a more aggregated measure of information. One function of the diversion matrix that one could require is that diversion from a product to all of its substitutes sums to one. This would not necessarily require knowing each individual element of a diversion matrix, but it would place a condition on aggregate diversion (i.e., that aggregate diversion, plus diversion to the outside good, sums to one). All discrete-choice models of demand assume that aggregate diversion plus diversion to the outside good sums to one, so this is a common, if sometimes implicit, assumption about the elements of a diversion matrix (or functions of its elements) that are important to measure.

When considering how to measure diversion, there are several potential methods. One approach is to estimate demand derivatives with a parametric model of demand, and measure diversion as the ratio of the estimated cross- and own-price derivatives. Another approach is to use second-choice data (e.g., changes to sales of substitutes when a product is removed, or responses to survey questions that identify a buyer's second-choice product) to measure diversion. The effects of interest are likely to vary across markets, leading the researcher or anti-trust authority to prefer different way of measuring diversion in different settings. 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 diversion ratio measures the outcome of this treatment, (i.e., the fraction of consumers who switch from j to a substitute product k). The treated group consists of consumers who would have purchased j at pre-existing prices, but no longer purchase j at a higher price.

[3] This differs from the 'diversion ratio matrix' of Jaffe and Weyl (2013). They extend the notion of upward pricing pressure to incorporate price responses by non-merging firms, and define the ratio of the quantity gained by a former rival's products to that lost by one's own products in response to a change in strategy in a matrix sense, holding fixed the strategy of the merger partner, but allowing all other strategies to adjust according to equilibrium play.

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.[4] A challenge of directly implementing such an experiment is that treating a small number of the most price-sensitive individuals lacks statistical power. An alternative is to treat all individuals who would have purchased j at pre-existing prices, and thus estimate an average treatment effect (ATE). 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 and MTE may diverge.

We illustrate the importance of choosing (i) which elements or functions of a diversion matrix to measure and (ii) the method of measuring diversion, in two applications. In the first application, we impose the assumption that aggregate diversion plus diversion to the outside good sums to one by estimating two discrete-choice models of demand. We use data from Nevo (2000) to explore the importance of different measures of diversion. We provide estimates of three different measures: a MTE, evaluated at pre-merger prices using a random-coefficients 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. We show that the ATE and MTE measures differ by 2-3% on average for each product's closest substitute, and by about $6-8\%$ for 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 MTE or ATE.

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; specifically, a set of vending machines. The experimental setting precludes us from estimating diversion that would be relevant to a small price change, but it does not require any restrictions on aggregate diversion. We find that even with a large-scale, appropriately randomized experiment, it is still necessary to control for variability in the overall level of demand. In order to control for unobserved demand shocks, we provide two conditions on economic

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

primitives and examine how they help to estimate experimental measures of the diversion ratio. The conditions are: (1) product removals cannot increase total sales, nor decrease total sales by more than the sales of the product removed, (2) that diversion to any single product is between 0 and 100 percent. We construct a non-parametric Bayesian shrinkage estimator to incorporate the second condition. We find that these two conditions improve our estimates of diversion, but that our estimates are sensitive to the strength of the prior. Next, we impose a different second condition: that aggregate diversion plus diversion to the outside good sums to one. This condition makes use of additional information in the market. We adapt our Bayesian shrinkage estimator, which allows us to nest both parametric structural estimates of diversion and (quasi)-experimental measures in a single framework, to incorporate this alternate condition. We find that even a very weak prior on aggregate diversion significantly improves our estimates.

We conclude that aggregate diversion is likely to be important for understanding diversion in a wide range of merger applications. It constrains estimates of diversion derived through discrete-choice models of demand, and improves estimates of diversion derived from a large-scale experiment. Our applications 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 in larger markets.[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 different measures of diversion can be obtained, how these measures of diversion might vary, and how to design and conduct experiments to measure diversion.

Finally, we consider several hypothetical mergers within the single-serving snack foods industry using the data from our second application. [SENTENCE ABOUT DIVESTITURE.]

1.1 Related Literature

There has been a recent debate on the use of experimental or quasi-experimental techniques vis-a-vis structural methods within industrial organization (IO) broadly, and within merger evaluation specifically. Angrist and Pischke (2010) complain about the general lack of experimental or quasi-experimental variation in many IO papers, and advocate viewing a merger itself as the treatment effect of interest. They cite Hastings (2004) as an example, which

[5]Examples include airlines (NEED CASE), Anhauser-Busch/InBev (NEED INFO).

[6]Examples include airline routes (NEED CASE), a merger of two European machinery firms, and data storage products in the Dell-EMC merger.

examines the effect of a merger between Thrifty and ARCO gasoline stations on the prices of competitors near and far away from affected stations. Nevo and Whinston (2010) 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, especially when both the proposal and approval of mergers is endogenous. They point out that while labor economics is often concerned with a single treatment effect, many of the key issues in IO are concerned with testing the equilibrium implications of economic theory, and that the complex competitive responses that arise in market settings often do not map well into the treatment effects framework.[7]

A deeper question, posed by 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. 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 growing debate about the role of unilateral effects has developed since the release of the 2010 Horizontal Merger Guidelines including: Carlton (2010), Schmalensee (2009), Willig (2011), and Hausman (2010).[8] There have also been recent attempts to validate the predictions of the unilateral effects approach in simulation (Miller, Remer, Ryan, and Sheu 2012) and empirically (Cheung 2011). Those papers find that the 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.

In spirit, our approach is most similar to Angrist, Graddy, and Imbens (2000), which shows how a cost shock can identify a particular local average treatment effect (LATE) for the price elasticity in a single product setting. That approach does not extend to the differentiated products setting because the requisite monotonicity condition may no longer be satisfied. We illustrate 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).

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

[8]Jaffe and Weyl (2013) incorporate an estimated pass-through rate to map anticipated opportunity cost effects into price effects.

Finally, by exploring the assumptions required for a credible (quasi)-experimental method of measuring diversion, we connect directly to the nascent theoretical literature discussing the use and measurement of the diversion ratio.[9] Farrell and Shapiro (2010) suggest that firms themselves track diversion in their 'normal course of business,' or that the diversion ratio is essentially another piece of data likely to be uncovered in a Hart-Scott-Rodino filing. Hausman (2010) argues that the only acceptable way to measure a diversion ratio is as the output from a structural demand system. Reynolds and Walters (2008) examine the use of stated-preference consumer surveys in the UK for measuring diversion.

The paper proceeds as follows. Section 2 provides a theoretical framework that motivates the use of a diversion ratio for merger analyses, and section 3 illustrates three different ways of measuring diversion within the context of a discrete-choice model of demand that imposes an assumption that aggregate diversion plus diversion to the outside good sums to one. Sections 4.3 and 4 illustrate three measures of diversion in the context of an experimental setting in the snack foods industry, in which we can measure diversion outside of the context of discrete-choice demand models. We discuss our experimental design, present results on three different ways of measuring diversion, and discuss the role that aggregate diversion plays in our estimates. Section 5 concludes.

2 Theoretical Framework

Merger guidelines focus on diversion as a key input for understanding the potential impact of a merger on prices. The rationale for this focus comes from an underlying supply-side model in which firms produce differentiated goods and compete according to a Nash equilibrium in prices. Farrell and Shapiro (2010) present these results to motivate the key constructs of the 2010 U.S. merger guidelines, and we review them here, using slightly different notation.

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

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

[9] 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).

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

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

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

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

Henceforth, we denote pre-merger costs as c_j and post-merger costs as $(1 - e_j) \cdot c_j$. A merger modifies the FOC of a single-product firm that owns j and acquires product k, when prices of all other goods p_{-j} are held fixed at the pre-merger values, to:

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

The extra term in the firm's post-merger FOC captures the degree to which a merger changes the opportunity cost of selling good j. After the merger, the firm internalizes the fact that some lost sales of product j are recaptured by product k. We denote the fraction of consumers who switch from j to k when faced with an increase in the price of j from the pre-merger prices $(p_j^{(0)}, p_{-j})$, holding all other prices fixed, as the diversion ratio $D_{jk}(p_j, p_{-j})$.

2.1 A Matrix of Diversion Ratios

Although the notion of diversion above is motivated by a differentiated-products Bertrand pricing equilibrium on the part of the acquiring firm in a merger, it sets aside the potential for rivals to respond strategically to a price increase.[10] This is an obvious area of interest for both researchers (e.g., Jaffe and Weyl and other citations) and practitioners. In the present work, however, we also restrict our focus to measuring diversion without attempting

[10]Farrell and Shapiro (2010) discuss some advantages of this restriction from the perspective of an antitrust practitioner, which include lower data requirements and the ability to avoid defining the relevant market.

to characterize equilibrium supply-side responses. Instead, we are interested in whether it may be important to consider whether diversion from product j to potential substitutes other than k might play a role in how well we are able to measure diversion from j to k. This is a question one can ask even after ruling out strategic responses from rivals. The answer may be illustrated by considering a matrix of diversion ratios:

$$\begin{bmatrix} (1 - D_{12} - D_{13}) & D_{12} & D_{13} \\ D_{21} & (1 - D_{21} - D_{23}) & D_{23} \\ D_{31} & D_{32} & (1 - D_{31} - D_{32}) \end{bmatrix}$$

where the diagonal elements of the matrix denote diversion to the outside good, and summing the off-diagonal elements of each row give aggregate diversion.

We illustrate with a hypothetical example. Suppose we can measure diversion from a Toyota Prius to three alternatives (a Honda Civic, a Tesla, and an outside good), writing this in the form of a matrix of diversion ratios:

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

In the example, aggregate diversion is 50 for a Civic, 70 for a Prius, and 80 for a Tesla. For the purpose of recovering an estimate of diversion between a Prius and a Civic, which elements of the matrix should we measure? In other words, if we're evaluating a merger between Toyota and Honda, is it sufficient to restrict our attention to the diversion from Prius to Civic, or is it important to understand other elements or functions of the matrix (e.g., aggregate diversion for the Prius)?

2.2 Diversion as a Treatment Effect

Once we determine the elements or functions of the diversion matrix that are necessary for evaluating the potential price effects of a merger, we can focus on the task of measuring each of the relevant elements, $D_{jk}(p_j, p^0_{-j})$. We define p_j as the post-merger price after an increase of Δp_j, so that $p_j = p^0_j + \Delta p_j$. We can then 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:

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

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 (1), we can re-write the diversion ratio using the marginal treatment effects (MTE) framework of Heckman and Vytlacil (2005), in which $D_{jk}(p_j, p_{-j}^0)$ is a marginal treatment effect that depends on p_j.[11]

$$\widehat{D_{jk}^{LATE}}(\Delta p_j) = \frac{1}{\Delta q_j} \int_{p_j^0}^{p_j^0+\Delta p_j} \underbrace{\frac{\partial q_k(p_j, p_{-j}^0)}{\partial q_j}}_{\equiv D_{jk}(p_j, p_{-j}^0)} \frac{\partial q_j(p_j, p_{-j}^0)}{\partial p_j}\, \partial\, p_j \quad (2)$$

$$\widehat{D_{jk}^{ATE}} = \frac{1}{\Delta q_j} \int_{p_j^0}^{\overline{p}_j} D_{jk}(p_j, p_{-j}^0) \frac{\partial q_j(p_j, p_{-j}^0)}{\partial p_j}\, \partial\, p_j = \left|\frac{q_k(\overline{p}_j, p_{-j}^0) - q_k(p_j^0, p_{-j}^0)}{q_j(\overline{p}_j, p_{-j}^0) - q_j(p_j^0, p_{-j}^0)}\right| \quad (3)$$

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 because the set of treated individuals varies with the size of the price increase. In equation (3) the average treatment effect (ATE) is the expression for the LATE where 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 infinitessimally small.[12] As we choose larger values for Δp_j our LATE estimate may differ from the MTE evaluated at $\mathbf{p^0}$.

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

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

[12] Anti-trust 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%.

infinitesimal" increase in price $\Delta p_j \gg 0$. We derive an expression for the second-order expansion of demand at (p_j, p_{-j}):

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

This allows us to compute an expression for the bias of a LATE estimate $\widehat{D_{jk}^{LATE}}(\Delta p_j)$ compared to the marginal diversion ratio $D_{jk}(p_j, p_{-j})$, as well as the variance of the LATE estimator (under the assumption of (locally) constant diversion, for which $\Delta q_k \approx D_{jk}\Delta q_j$):

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

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

The expression in (5) shows that the bias depends on two things: one is the magnitude of the price increase Δp_j, the second is the curvature of demand (the terms $\frac{\partial^2 q_j}{\partial p_j^2}$ and $\frac{\partial^2 q_k}{\partial p_j^2}$). This suggests that bias is minimized by considering small price changes. The disadvantage of considering a small price change Δp_j is that it implies that the size of the treated group Δq_j is also small, and thus the variance of our diversion measure is large, as shown in equation 6. 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 potential variance; a larger Δp_j (and by construction Δq_j) may yield a less noisy LATE, but may deviate from the quantity of interest.

A relevant question is: What are the economic implications of assuming a constant treatment effect, such that $D_{jk}(p_j, p_{-j}) = D_{jk}$? We can see the answer by examining the case where the bias calculation in equation (5) is equal to zero. Economic theory provides guidance because the underlying objects are demand curves. Two functional forms for demand exhibit constant diversion and are always unbiased: the first is linear demand, for which $\frac{\partial^2 q_k}{\partial p_j^2} = 0$, $\forall j, k$. The second is the IIA logit model, for which $D_{jk} = -\frac{\partial^2 q_k}{\partial p_j^2} / \frac{\partial^2 q_j}{\partial p_j^2}$. *Implicitly when we assume that the diversion ratio does not vary with price, we assume that the true demand system is well approximated by either a linear demand curve or the IIA logit model.* We derive these relationships, as well as expressions for diversion under other demand models in Appendix A.1, and show that random coefficients logit demand, and CES demands (including log-linear demand) do not generally exhibit constant diversion.

If the primary concern in a given market is that the curvature of demand is steep, so that assuming a constant diversion ratio is unreasonable, one may need to consider a small price increase to avoid bias. However, if the primary concern is that sales are highly variable, one may need to consider a larger price increase to reduce variance.[13] Information about the elasticity (and super-elasticity) of demand for j can be informative. As noted in (2), the LATE estimate will concentrate more weight near $\mathbf{p^0}$ when demand is more elastic, or if demand becomes increasingly inelastic as Δp_j becomes larger.[14]

In figure 1, we consider what diversion might look like for three different hypothetical demand curves for a Toyota Prius. In the first example, we consider diversion to three alternatives (a Honda Civic, Tesla, and an outside good) when demand for a Prius is linear. As we vary price upwards from $25,000 to $50,000, diversion to the three alternatives is constant (as is required by linear demand): 63% of potential Prius buyers switch to a Honda Civic, 12% switch to a Tesla, and 25% switch 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 to the same alternatives when demand for a Prius has a constant elasticity of substitution (CES) of -1. We refer to this as an inelastic CES demand curve. 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, and more substitution to the Tesla after a large price increase. Using the histogram to weight the diversion across the entire price spectrum (from $25,000 to $50,000) provides an estimate of aggregate diversion (ATE) that is 59% to the Honda Civic, 18% to the Tesla, and 22% to the outside good. However, for a small price increase, diversion to the Honda Civic is over 90%.

The third, and final example in figure 1 considers diversion when demand for a Prius is CES, but with an elasticity of -4. This greater elasticity changes the relative weighting

[13]Furthermore, if the policy maker believes that a proposed merger could induce firms to withdraw a product from the market or impose a large price increase, then the ATE calculation becomes the primary object of interest.

[14]In our empirical example, it might seem reasonable that customers who substitute away from a *Snickers* bar after a five cent price increase switch to *Reese's Peanut Butter Cup* at the same rate as after a 25 cent price increase, where the only difference is the number of overall consumers leaving *Snickers*. However in a different industry, this may no longer seem as reasonable.

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 CES demand, the aggregate diversion (ATE) is now more heavily weighted towards consumers that leave at small price changes. Thus, aggregate diversion to the Honda Civic is 72%, with 10% diverting to a Tesla and 18% diverting to the outside good.

To summarize, the LATE/ATE provides a good approximation for the MTE at $\mathbf{p^0}$ when the bias in (5) is small, which happens: (a) when the curvature of demand is low ($\frac{\partial^2 q_k}{\partial p_j^2} \approx 0$), (b) when the true diversion ratio is constant (or nearly constant) so that $D_{jk}(p_j, p_{-j}) = D_{jk}$, or (c) when demand for j is steepest near the market price $\left|\frac{\partial q_j(p_j, p_{-j})}{\partial p_j}\right| \gg \left|\frac{\partial q_j(p_j+\Delta p_j, p_{-j})}{\partial p_j}\right|$.

2.3 Second-Choice Data

Often researchers have access to one form or another of 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. The UK Competition Commission makes use of consumer surveys as part of the merger review process both for market definition, and for second-choice data (Reynolds and Walters 2008). The FCC and DOJ made use of observed consumer switching data in the analysis of the proposed AT&T/T-Mobile merger.

Exploiting variation in consumer choice sets provides a revealed-preference mechanism for recovering second-choice data. A problem with using observational variation in choice sets is that the variation is often non-random. If one simply compares 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.[15]

A more direct approach is to construct second-choice data experimentally by removing product j from a consumer's choice set for a period of time. One way to interpret second-choice data or an experimental product removal is as an increase in price to the choke price $\overline{p}_j$, where $q_j(\overline{p}_j, p_{-j}) = 0$. One advantage of the product removal experiment is that it treats as many individuals as possible, and thus minimizes the variance expression in (6). The

[15]In previous 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.

downside is that it measures the ATE, which may or may not be the quantity of interest, depending on the policy questions most critical to evaluating the merger.

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:

$$\widehat{D_{jk}^{ATE}} = \left| \frac{q_k(\overline{p}_j, p_{-j}^0) - q_k(p_j^0, p_{-j}^0)}{q_j(\overline{p}_j, p_{-j}^0) - q_j(p_j^0, p_{-j}^0)} \right| = \frac{q_k(\mathbf{p^0}, A \setminus j) - q_k(\mathbf{p^0}, A)}{q_j(\mathbf{p^0}, A)} \tag{7}$$

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. Suppose that instead of increasing the price p_j, we reduced the quality ξ_j of good j. This would trace out a different LATE of the most quality-responsive consumers, rather than the most price-responsive consumers:

$$\widehat{D_{jk}^{LATE,\xi}}(\Delta \xi_j) = \int_{\xi_j^0}^{\xi_j^0 - \Delta \xi_j} \left| \frac{\frac{\partial q_k(\mathbf{p^0}, \xi_j)}{\partial \xi_j}}{\frac{\partial q_j(\mathbf{p^0}, \xi_j)}{\partial \xi_j}} \right| d\,\xi_j \tag{8}$$

However, if we consider a large enough decrease in quality $\xi_j \to -\infty$ so that no individual purchases good j, then $\widehat{D_{jk}^{LATE,\xi}}(\infty) \to \widehat{D_{jk}^{ATE}}$ (which no longer depends on the instrument). This highlights two important points: the ATE may be identified even if $\frac{\partial q_k}{\partial p_j}$ is not, and the ATE estimates may not be subject to the same endogeneity concerns as the price effects.

3 Application to Nevo (2000)

In our first application, we use the well-known example from Nevo (2000). The data consist of $T = 94$ markets with $J = 24$ brands per market and a $I = 20$ point distribution of heterogeneity for each market. Recall, the specification allows for product fixed effects d_j and both 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).[16] One important feature of this example is that it features a relatively large amount of preference heterogeneity, especially with respect to the price sensitivity β_{it}^{price}. The estimated coefficient on price is distributed as follows:

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

We denote a measure of diversion evaluated for an infinitessimally small price change as MTE_p, and a measure of diversion evaluated for an infinitessimally small change in quality as MTE_q. We refer to a 'second choice' estimate of diversion as an ATE. For comparison, we also evaluated a Logit model, under which diversion is assumed to be constant. These four treatment effects are defined as:

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

We suppose that the policy-relevant calculation of interest is the MTE_p, and we quantify how well the other treatment effect estimators approximate MTE_p. We compare MTE_p to: MTE_q (the marginal diversion ratio calculated by reducing the quality $\Delta\xi_{jt}$ of good j rather than increasing its price); ATE (the average treatment effect that we would identify from a product removal or second choice data); and $Logit$ (the diversion ratio we would estimate if we assumed that diversion was proportional to the marketshares as in the IIA logit model).

The first comparison we look at is for each of the 94 markets and 24 products to compute the best substitute for each product-market pair, and to calculate the diversion ratio to that product. In Table 1, we report these patterns. We find that for MTE_p, MTE_q 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 MTE_q and ATE identify the same best substitute as the MTE_p around 90% of the time, while the plain logit (which identifies the same best substitute for all products) is only in agreement 58% of the time. We can repeat the exercise and calculate substitution to the outside good. Here we find that MTE_p has slightly more outside good substitution $(35 - 37\%)$ than the

[16]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 identical point estimates.

	MTE_p	MTE_q	ATE	*Logit*
		Best Substitute		
Med(D_{jk})	13.26	13.45	13.54	9.05
Mean(D_{jk})	15.11	15.43	15.62	10.04
% Agree with MTE_p	100.00	89.80	89.98	58.38
		Outside Good		
Med(D_{j0})	35.30	32.59	32.40	54.43
Mean(D_{j0})	36.90	33.90	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.

MTE_q or ATE diversion measures ($32 - 33\%$), 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. We treat MTE_p as the 'true' value and compare the difference in the calculated diversion. For example, we compare the log difference between $\log D_{jk}^{ATE} - \log D_{jk}^{MTE_p}$.[17] 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. As indicated in Table 2, this implies that MTE_q and ATE measures of diversion are on average $2 - 3\%$ higher than the MTE_p measure of diversion for each product's best substitute. Across all substitute products, shown in the second panel, the ATE and MTE_q measures are around $6 - 8\%$ higher than the MTE_p measure. When compared to the outside good, the ATE and MTE_q measures are around $8-9\%$ lower than the MTE_p 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 as much as 39%.

There is no obvious theoretical reason as to why the ATE and MTE_q measures would overstate (understate) substitution to other products on average when compared to the MTE_p measure, other than the fact that the marginal consumer tends to become more

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

	med($y-x$)	mean($y-x$)	med($\|y-x\|$)	mean($\|y-x\|$)	std($\|y-x\|$)
			Best Substitutes		
MTE_q	1.79	2.36	5.82	7.29	6.46
ATE	2.56	3.24	6.00	7.61	7.04
Logit	-44.19	-42.88	44.92	47.77	28.63
			All Products		
MTE_q	5.65	8.40	8.17	12.14	12.27
ATE	5.78	8.30	8.29	12.13	12.02
Logit	-35.90	-25.92	49.48	53.27	34.56
			Outside Good		
MTE_q	-7.46	-8.48	7.46	8.66	6.64
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 $x = \log(\widehat{D^{MTE_p}})$
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_p 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_p diversion to the outside good.

(less) inelastic as the price increases due to the curvature of demand induced by the logit error term.[18] What does appear to be clear is that when we reduce an estimator's ability to accommodate heterogeneity in consumer preferences, the MTE_p, MTE_q, and ATE measures get closer together. In Appendix A.2, we repeat this exercise with a restricted version of the demand model at the original published estimates from Nevo (2000).[19] In Appendix A.3 we conduct Monte Carlo simulations with commonly-used parametric demand models, and report the maximum discrepancy between the MTE and ATE estimates.

4 Empirical Application to Vending

In our next empirical application, we estimate the ATE form of the diversion ratio using observed second-choice data. We run a series of experiments in which we exogenously remove a product from 66 vending machines located in office buildings in Chicago, and measure substitution to the remaining products. We begin with a discussion of the snack foods/vending

[18]The (random coefficients) logit model has an inflection point at $s = 0.5$. At the market level we know that $s_j < 0.5$ for all j except for the outside good. We have $s_0 > 0.5$ in some markets and $s_0 < 0.5$ in others. At the individual level, it is not uncommon for $s_{ij} > 0.5$. For these reasons, we cannot necessarily sign the second derivative of demand $\frac{\partial^2 q_k}{\partial p_j^2}$.

[19]The restriction imposed is that $\pi_{inc^2,price} = 0$.

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.

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*.[20] We report HHI's at both the category level and for all vending products in Table 3 from the region of the U.S. that includes Chicago. If the relevant market is defined at the category level, all categories are considered highly concentrated, with HHIs in the range of roughly 4500-6300. If the relevant market is defined as all products sold in a snack-food vending machine, the HHI is below the critical threshold of 2500. Any evaluation of a merger in this industry would hinge on the closeness of competition, and thus require measuring diversion.

Over the last 25 years, the industry has been characterized by a large amount of merger and acquisition activity, both on the level of individual brands and entire firms. For example, the *Famous Amos* cookie brand was owned by at least seven firms between 1985 and 2001, including the Keebler Cookie Company (acquired by Kellogg in 2001), and the Presidential Baking Company (acquired by Keebler in 1998). *Zoo Animal Crackers* have a similarly complicated history, having been owned by Austin Quality Foods before they too were acquired by the Keebler Cookie Co. (which in turn was acquired by Kellogg).[21]

[20]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.

[21]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.]

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 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.[22] 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.[23]

In addition to the data from Mark Vend, we also collect data on the characteristics of each product online and through industry trade sources.[24] For each product, we note its manufacturer, as well as the following set of product characteristics: package size, number of servings, and nutritional information.[25]

4.2 Experimental Design

We ran four exogenous product removals with the help of Mark Vend Company. These represent a subset of a larger group of eight exogenous product removals that we have analyzed in two other projects, Conlon and Mortimer (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

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

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

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

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

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.[26] These machines were also located on routes that were staffed by experienced drivers, which maximized the chance that the product removal would be successfully implemented. The 66 machines used for each treatment are distributed across five of Mark Vend's clients, which had between 3 and 21 machines each. The largest client had two sets of floors serviced on different days, and we divided this client into two sites. Generally, each site is spread across multiple floors in a single high-rise office building, with machines located on each floor.

For each treatment, we remove a product from all machines at a client site for a period of 2.5 to 3 weeks. The four products that we remove are the two best-selling products from either (a) chocolate maker Mars Incorporated (Snickers and Peanut M&Ms) or (b) cookie maker Kellogg's (Famous Amos Chocolate Chip Cookies and Zoo Animal Crackers). We refer to exogenously-removed products as the *focal products* throughout our analysis.[27] Whenever a product was exogenously removed, poster-card announcements were placed at the front of the empty product column. The announcements read: *This product is temporarily unavailable. We apologize for any inconvenience.* The purpose of the card was two-fold: first, we wanted to avoid dynamic effects on sales as much as possible, and second, Mark Vend wanted to minimize the number of phone calls received in response to the stock-out events.

The dates of the interventions range from June 2007 to September 2008, with all removals run during the months of May - October. We collected data for all machines for just over three years, from January of 2006 until February of 2009. During each 2-3 week experimental period, most machines receive service visits about three times. However, the length of service visits varies across machines, with some machines visited more frequently than others. Though 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. In order to do this, we assume that sales are distributed uniformly among the business days in a service

[26] 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.

[27] Not reported here are two experiments on best-selling products from Pepsi's Frito Lay Division, which we omit for space considerations, and because Pepsi's products already dominate the salty snack category (which makes merger analysis less relevant). We also ran two additional experiments in which we removed two products at once; again we omit those for space considerations and because they don't speak to our diversion ratio example. These are analyzed in Conlon and Mortimer (2013b) and Conlon and Mortimer (2017).

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.[28]

The cost of the experiment consisted primarily of driver costs. Drivers had to spend extra time removing and reintroducing products to machines, and the driver dispatcher had to spend time instructing the drivers, tracking the dates of each experiment, and reviewing the data as they were collected. Drivers are generally paid a small commission on the sales on their routes, so if sales levels fell dramatically as a result of the experiments, their commissions could be affected. Tracking commissions and extra minutes on each route for each driver would have been prohibitively expensive to do, and so drivers were provided with $25 gift cards for gasoline during each week in which a product was removed on their route to compensate them for the extra time and the potential for lower commissions.

Our experiment differs somewhat from an ideal experiment. Ideally, we would be able to randomize the choice set on an individual level. Technologically, of course, that is difficult in both vending and traditional brick and mortar contexts. In contrast, online retailers are capable of showing consumers different sets of products and prices simultaneously. This leaves our design susceptible to contamination if for example, Kraft runs a large advertising campaign for Planters Peanuts that corresponds to the timing of one of our experiments. Additionally, because we remove all of the products at an entire client site for a period of 2.5 to 3 weeks, we lack a contemporaneous "same-side" group of untreated machines. We chose this design, rather than randomly staggering the product removals, because we (and the participating clients) were afraid consumers might travel from floor to floor searching for stocked-out products. This design consideration prevents us from using 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.

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

4.3 Description of Experimental Data

We summarize the data generated by our experiments 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. 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) [29]

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.

A second challenge is that there is a large amount of variation in overall sales at the weekly level, independent of our product removals. This weekly variation in overall sales is common in many retail environments. We often observe week-over-week sales that vary by over 20%. This can be seen in Figure 2, which plots the overall sales of all machines from one of the sites in our sample on a weekly basis. In our particular setting, many of the product removals were implemented during the summer of 2007, which was a high-point in demand at several sites, most likely due to macroeconomic conditions.

We explore this relationship further in Table 5, where we report the average sales by week during both the treatment and control periods for key substitutes. The third column reports the quantile that the sales during the mean treatment week corresponds to in the distribution of control weeks. For example, during our Snicker's removal experiment 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

[29]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 experiments 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.

compared to a control average of 4,892 which corresponded to the 74.4th percentile of the control distribution for total sales.

4.4 A Simple Matching Estimator for Diversion

There are a number of challenges to constructing an estimate of the diversion ratio: (1) the overall size of the market/rate of consumer arrivals varies substantially over time including among treatment and control periods; (2) the set of substitute products may vary across locations/machines; and (3) holding fixed the market size, variation in product-level sales could imply that $\Delta q_k < 0$ or $\Delta q_k > |\Delta q_j|$.

Consider the Wald-type estimator, where $Z = 1$ denotes the removal of product j:[30]

$$\widehat{D_{jk}} = \frac{\widehat{\Delta q_k}}{|\widehat{\Delta q_j}|} = \frac{E[q_k|Z=1] - E[q_k|Z=0]}{|E[q_j|Z=1] - E[q_j|Z=0]|}$$

We want to adjust our calculation of the expectation to address problem (1) as documented in Table 5. In other words, we want to adjust our control group to account for the fact that (on average) the treated weeks represent the 74th percentile (rather than the mean) of the overall sales distribution. To be explicit about the problem, we introduce a covariate x (demand shock):

$$E[q_k|Z=z] = \int q_k(x,z) f(x|Z=z)\, d\, x$$

The problem is that $f(x|Z = 1) \neq f(x|Z = 0)$, the treated and control periods have different distributions of covariates (demand shocks). The typical solution involves *matching* or *balancing*, where one re-weights observations in the control period using measure $g(\cdot)$ so that $f(x|Z = 1) = g(x|Z = 0)$ and then calculates the expectation $E_g[q_k|Z = 0]$ with respect to measure g.[31] For each treated week t, we can construct a set of matched control weeks within a neighborhood $S(x_t)$:

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

[30] One advantage of using product removal experiments is that $E[q_j|Z = 1] = 0$ 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.

[31] We omit discussion of the conditional independence assumption because we have randomized assignment of Z.

Our first assumption places some obvious restrictions on potential control weeks:[32]

Assumption 1. *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.*

The remaining question is how to define $S(x_t)$, the neighborhood of "similar enough" matches. Abadie and Imbens (2006) consider k-nearest neighbors in the space of x_t.[33] Our design is complicated by the fact that we don't directly observe the demand shock x_t. Instead, we derive weaker conditions that help to balance the treatment and control periods without observing x_t directly. Our next assumption is a weak implication of all products being substitutes for one another.

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.*

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:

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

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 $\widehat{q_{js}} = E[q_{js}|Q_s, Z = 0]$. 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:

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

[32] The first assumption is required because otherwise q_{ks} is not defined. The second is designed to prevent our procedure from introducing additional unobserved heterogeneity. By restricting potential controls to different weeks at the same machine, we attempt to control for unobserved machine-level heterogeneity.

[33] 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.

Thus (11) defines the set of control periods S_t which correspond to treatment period t under our assumptions. The economic implication of Assumption 2 is that the sum of the diversion ratios from j to all other products is between zero and one (for each t): $\sum_{k \neq j} D_{jk,t} \in [0, 100\%]$.

We explore the implications of Assumptions 1 and 2 and report our estimates of $\widehat{\Delta q_j}, \widehat{\Delta q_k}$ and $\widehat{D_{jk}} = \frac{\widehat{\Delta q_k}}{\widehat{\Delta q_j}}$ in Table 6 for the Snickers removal experiment.[34] The table is broken up into two panes. Each pane displays the number of treated weeks for each substitute, the average number of control-weeks matched to each treatment week, and our estimates for $\Delta q_j, \Delta q_k$ and $D_{jk} = \frac{\Delta q_j}{\Delta q_k}$. In both panes, there is a large amount of variation in the number of treatment-weeks (and hence focal product sales: Δq_j) across substitute products. The main source of this variation is that not every machine stocks every product, for some substitutes we have over 180 treated machine-weeks, while for others we have fewer than 10. With the addition of Assumption 2, we reduce the # of control weeks per treated machine-week from 90-120 to 7-10. In some cases, Assumption 2 eliminates treated machine-weeks because it cannot find any matches though this effect is generally small (less than 10% of treated machine-weeks).

The addition of Assumption 2 also improves our estimates of the diversion ratio $\widehat{D_{jk}} = \frac{\widehat{\Delta q_k}}{\widehat{\Delta q_j}}$ somewhat. For example, with only Assumption 1 diversion to the outside good is measured as $D_{j0} = \frac{-982}{|-929|} = -106\%$, with the addition of Assumption 2 this constrains the set of potential control weeks and we estimate $D_{k0} = \frac{461}{|-920|} = 47.5\%$. However, the matching estimator alone does not produce "reasonable" looking estimates of the diversion ratio. Nearly half of products exhibit negative diversion ratios, and in both panes diversion to the top 5 substitutes exceeds 200%, so that two products are purchased for every lost Snickers sale. In both panes, the "best substitute" is the *Consolidated Non-Chocolate Nestle Candy* (Willy Wonka, Runts, etc.) and the second best substitute is *M&M Peanut*. M&M Peanut is available in nearly every machine and we estimate diversion ratios of 52.7% and 39.4% respectively; the treatment mean exceeds any of the weekly sales during the control period (recall Table 5). The *Consolidated Non-Chocolate Nestle Candy* is only stocked in two machines, and its "best substitute" status is driven by the small size of its sample $D_{jk} = \frac{9.4}{|-10.5|} = 89.5\%$.

We were able to address problem (1), variation in the overall level of demand. We were able to partially address problem (2) by restricting our sample to cases where the substitute

[34] For our other removals: M&M Peanut, Famous Amos Cookies, and Animal Crackers consult the Online Appendix.

k was available, though this lead to differentially sized treatment groups Δq_j for different substitutes. The matching estimator did not do much to address problem (3), the noise in our estimates of the diversion ratio caused by variance at the product level; nor does it guarantee that our estimates of diversion look "sensible."

4.5 A Nonparametric Bayesian Estimator

Now we ask: given our matched estimates of $\widehat{\Delta q_j}$ and $\widehat{\Delta q_k}$ can we construct a better estimate of D_{jk} than $\frac{\widehat{\Delta q_k}}{\widehat{\Delta q_j}}$? At first pass, this may seem like a ridiculous idea. If $D_{jk} \equiv \frac{\Delta q_k}{\Delta q_j}$, how can there be a better estimate than $\frac{\widehat{\Delta q_k}}{\widehat{\Delta q_j}}$? However, we may be willing to place *ex-ante* restrictions on the domain of D_{jk}, for example we might be willing to rule out negative diversion ratios or diversion ratios in excess of 100%. Likewise, there may be additional information in pooling estimates of diversion ratios across substitute products k. We consider two additional assumptions below:

Assumption 3. *"Unit Interval": $D_{jk} \in [0, 1]$.*
$\Delta q_k | \Delta q_j, D_{jk} \sim Bin(n = \Delta q_j, p = D_{jk})$ and $D_{jk} | \mu_{jk}, m_{jk} \sim Beta(\mu_{jk}, m_{jk})$.

Assumption 4. *"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 $D_{jk} \sim Dirichlet(\mu_{j0}, \mu_{j1}, \ldots, \mu_{jK}, m_{jk})$.

Assumption 3 restricts D_{jk} to the unit interval, while Assumption 4 goes further and restricts the *vector* $\mathbf{D_{j\cdot}}$ to the unit simplex. In order to impose this structure we take a nonparametric Bayes approach. It is *nonparametric* in the sense that when we observe Δq_j trials of treated individuals and Δq_k successes and are looking to estimate the probability of success $p = D_{jk}$, the binomial distribution is not a substantive restriction. Under Assumption 4, the Dirichlet prior restricts the sum of the diversion ratios (for all goods including the outside good): $\sum_{\forall k} D_{jk} = 1$. It is *nonparametric* in the sense that when there are K substitutes, there are $K + 1$ data points and $K + 1$ free parameters.

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.[35] This formula makes it easy to express the posterior

[35] 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.

mean (under Assumption 3) as a *shrinkage estimator* which combines our prior information with our experimental data:

$$\widehat{D_{jk}} = \lambda \cdot \mu_{jk} + (1-\lambda)\frac{\Delta q_k}{\Delta q_j}, \quad \lambda = \frac{m_{jk}}{m_{jk} + \Delta q_j} \tag{12}$$

Here λ tells us how much weight to put on our prior mean versus our experimental observations, and directly depends on how many "pseudo-observations" we observed 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), $\hat{D}_{jk}$ is shrunken towards μ_{jk} (from either direction). Thus, when our experiments provide lots of information about diversion from j to k we rely on the experimental outcomes, but when our experiments are less informative we rely more on our prior information.[36]

Under either Assumption 3 or 4, the remaining challenge is how to select μ_{jk}. 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. The IIA logit prior is useful 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.[37] 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. Generally speaking, a smaller m implies a *weaker prior* and more weight on the observed data.

When μ_{jt} is chosen as a function of the same observed dataset (including from estimated demand parameters) this is a form of an *Empirical Bayes* estimator. The development of Empirical Bayes shrinkage is attributed to Morris (1983) and has been widely used in applied microeconomics to shrink outliers from a distribution of fixed effects in teacher value added[38]

[36]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 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 consult Appendix A.4.

[37]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.

[38]Chetty, Friedman, and Rockoff (2014) and Kane and Staiger (2008)

or hospital quality.[39]

4.6 Estimates of Diversion

For each of our four experiments, we report our estimates of the diversion ratios in Table 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 and strong prior), and the "multinomial" version of the diversion ratio under Assumption 4. When we incorporate a prior distribution, we center the mean at the IIA logit estimates $\mu_{jk} = \frac{s_k}{1-s_j}$.[40] 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 7, $\Delta q_k = 289.6$ 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.[41] 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 ($m_{jk} = 64$ pseudo-observations, one for each potential substitute), we shrink the estimates of both 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 $m_{jk} = 300$ 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 $m = 3.05$ 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

[39] Chandra, Finkelstein, Sacarny, and Syverson (2013)

[40] For the Dirichlet we add an additional small (uniform) $\frac{1.3}{K+1}$ term to the logit probabilities in order to bound some of the very small prior probabilities 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.

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

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 went from our "best" substitute, to hardly a substitute at all, while Twix remained the second best substitute (now behind M&M Peanut which had a similar "raw" diversion measure, but a larger treated group $\Delta q_j = -954.3$ compared to $\Delta q_j = -702.4$). The simplex restriction shrinks outside good diversion from 47.5 $\rightarrow$ 23.1.

In Table 9, we report the posterior distribution of our preferred diversion estimates under Assumption 4 and the very weak prior $m = 3.05$. 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.[42] On the other, it demonstrates the power of cross substitute restrictions in Assumption 4; even with a diffuse prior, and very little experimental data for some substitutes, the simplex is sufficient to pin down diversion ratios.

While Assumption 4 appears relatively innocuous (most researchers are likely willing to assume a multinomial discrete choice framework) because it is so powerful in pinning down the diversion ratio estimates, we should be a little cautious. The important empirical 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, we were interested in a merger where product j acquired both (k, k') but (k, k') were always rotated for one another and never available at the same time, we might want to vary the set of products over which we sum $D_{jk'}$ for each alternative: $\mathcal{K}_k$.[43]

One of the perceived benefits of the unilateral effects approach is that it requires data only from the merging parties, and not from firms outside the merger.[44] 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}.[45] This suggests that although we need only (quasi)-experimental removals (or second-choice data) for the focal products involved in the

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

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

[44]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.

[45]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).

merger, we should attempt to measure substitution to all available substitutes if possible.

4.7 Merger Evaluation

The goal behind estimating the diversion ratio is to allow regulators to perform prospective merger analysis. Within the unilateral-effects framework, the diversion ratio is the key input for calculating UPP or the GUPPI. These two measures also rely on estimates of price-cost margins, which firms are compelled to provide as part of the Hart-Scott-Rodino filings.

The GUPPI for a merger in which the seller of product j acquires product k is $GUPPI_j = \frac{p_k - c_k}{p_j} \cdot D_{jk}$ (all evaluated at pre-merger values). An estimate of UPP would further require some information (or an assumption) about the potential size of the variable cost reduction from the proposed merger. Under the assumption that $p_k - c_k$ is identical for all products sold by the acquired firm $\mathcal{F}_k$[46], we can rearrange the expression in UPP from (**??**) to find the *critical diversion ratio*:

$$UPP_j > 0 \iff \left(\sum_{k \in \mathcal{F}_k} D_{jk} \right) > \frac{e_j \cdot c_j}{p_k - c_k} \tag{13}$$

In our vending application, we observe both the retail prices and prices paid by retailers to the manufacturers; but we do not observe the marginal costs of the manufacturers. Thus, we would require additional assumptions in order to report UPP or the GUPPI. Although there is some variation in wholesale prices, most fall between 35 and 50 cents. In our previous work Conlon and Mortimer (2017), we found that $c = 0.15$ appeared to be a reasonable estimate of manufacturer marginal cost for confections. If we apply this cost (for all products) to a wholesale price of $p = 0.45$ then (13) simplifies to $\sum_{k \in \mathcal{F}_k} D_{jk} > \frac{1}{2} \cdot e_j$. In other words, we would require that the sum of the diversion ratios to all newly acquired substitutes be less than half as large as the anticipated marginal cost savings. Thus for a 10% reduction in marginal costs, we would require that the sum of diversion ratios be less than 5%.

An important remedy available to the antitrust agencies is that mergers can be approved conditional on some divestiture.[47] We can measure whether divesting a product during a

[46] Whether or not this assumption is reasonable is context specific. It is likely to be innocuous for many consumer products where different flavors or variants have similar wholesale prices and costs (e.g., the margin differences between blueberry and strawberry yogurt are likely insignificant). In other contexts, such as automobiles or airlines, margin differences are likely to be more important for understanding pricing effects.

[47] 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.

merger reduces $\sum_{k \in \mathcal{F}_k} D_{jk}$ below some critical threshold. In Table 10, we consider whether or not a merger would result in upward pressure on the prices of the products we experimentally removed. We then propose a divestiture of a key substitute product controlled by the target and recompute the diversion ratio to all of the target's products absent the divested product. In some circumstances, divestiture of one or two key products might alleviate upward pricing pressure concerns around a particular merger.

We examine a potential acquisition in which Kellogg's (Pop-Tarts, Zoo Animal Crackers, Famous Amos Cookies, Cheez-it, Rice Krispie Treats) acquires Kraft's snack food division (Oreos, Lorna Doone, Planters Peanuts, Cheese Nips, and other Nabisco products). We examine the effects on both of Kellogg's major products (Zoo Animal Crackers and Famous Amos) of the Kraft acquisition. We find that the diversion ratio from Zoo Animal Crackers to all Kraft products is 5.80%, and the diversion of Chocolate Chip Famous Amos is 11.85%. Both of these are above the 5% threshold corresponding to a 10% reduction in marginal costs, and thus would need to demonstrate substantial cost synergies to justify the merger. However, if Kraft were to divest its Planter's Peanut line, the diversion ratios drop to 3.36% and 3.09% respectively. If anticipated marginal cost savings were 6% or more, this suggests a potential remedy that might allow the antitrust authorities to drop opposition to the merger.

Likewise one could consider an acquisition by Mars (Snickers, M&M's, Milky Way, Three Musketeers, Skittles) of Nestle's US confections business (Butterfinger, Raisinets, assorted Willy Wonka fruit flavored candies). If one is worried about the price effects that the acquisition might have on Snickers or M&M Peanut (the two largest brands in the chocolate category) we find that diversion from those products to Nestle products is 5.71% and 6.30% respectively, but that if Nestle were to divest Butterfinger, the diversion ratios would drop to 1.26% and 4.51% respectively. Again, this might be enough to convince the antitrust authorities not to block a proposed acquisition.[48]

Our hope is that these examples highlight both the advantages of our approach (that it is easy to detect which mergers require further investigation and which divestitures to consider), but also some of the limitations. For example, we can look at the effect on Snickers or M&M Peanut that the acquisition of Butterfinger might have, but we cannot say anything about the likely effect on the prices of Butterfinger of a Mars acquisition without conducting that experiment as well. This suggests that we either need observational/quasi-experimental data on many different stockout events, or we need some *ex ante* idea of which products are likely

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

to have larger price impacts of the merger in order to tailor our experiments. The second limitation, which is not a limitation of our approach but of the unilateral effects approach more generally, is that it ignores diversion to existing brands. In the Snickers experiment, more than half of consumers already substitute to another Mars product, yet this has no bearing on the analysis of a proposed merger with Nestle (though it might if we considered the price effects on Butterfinger). This highlights what is likely to be a more general pattern in the unilateral effects approach: when large brands acquire smaller brands, the likely concern is the price of the smaller brand.

5 Conclusion

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

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

We conduct randomized field experiments, where we exogenously remove products from consumers' choice sets and measure the ATE directly. We provide a set of three relatively minimal assumptions, derived from consumer theory, which allow for relatively precise estimates of the diversion ratio even in noisy environments: (1) that a product removal cannot increase sales, nor decrease sales by more than the expected sales of the removed product (2) that the diversion ratio is defined on the unit interval and has some prior distribution and (3) that the diversion ratio for all possible substitutes is defined on the unit simplex. We find benefits from measuring diversion not only between products involved in a proposed merger, but also from merging products to non-merging products.

We develop a simple method to recover the diversion ratio from data, which enables us to combine both experimental and quasi-experimental measures with structural estimates as prior information. 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. Structural demand estimates rely either on variation in the prices of choices across contexts or markets, or on the set of choices that are available to consumers in order to identify parameters. These estimators struggle when there is little variation in the data. In contrast, randomized controlled trials work best when there is little to no variation in the availability or prices of products, except for the variation induced by the treatment.

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

References

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

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

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

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

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

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

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

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

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

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

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

Conlon, C. (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.

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.

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

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

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

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

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

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

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

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

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

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

Nevo, A. (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.

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.

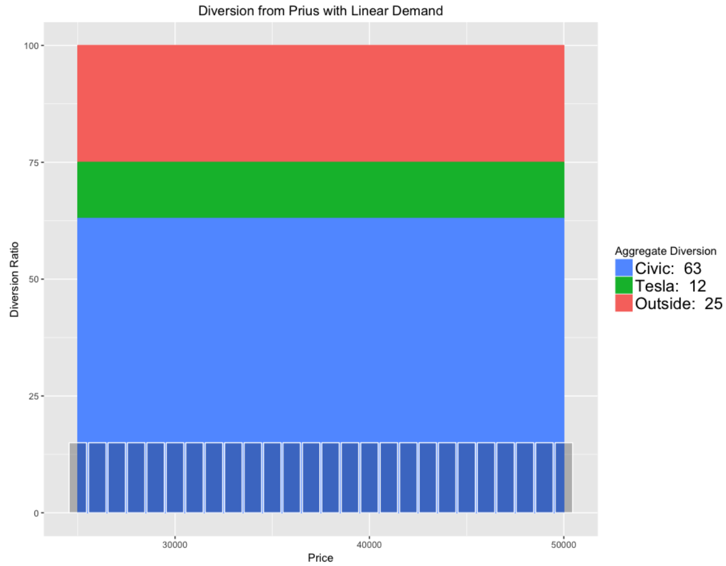
Team, S. (2015): "RStan: the R interface to Stan, Version 2.8.0," .

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.

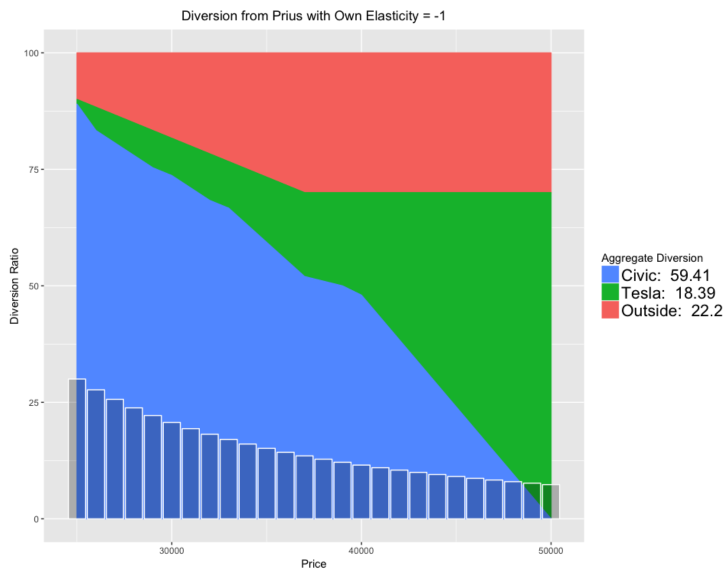
Werden, G. J., and L. Froeb (2006): "Unilateral competitive effects of horizontal mergers," *Handbook of AntitrustEconomics.*

Whinston, M. D. (2006): *Lectures on Antitrust Economics.* MIT Press.

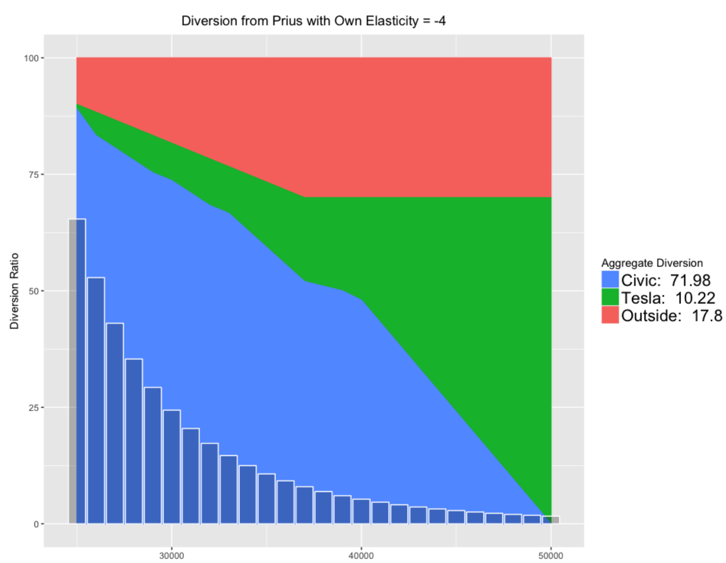
Willig, R. (2011): "Unilateral Competitive Effects of Mergers: Upward Pricing Pressure, Product Quality, and Other Extensions," *Review of Industrial Organization*, 39(1-2), 19–38.



(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

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

Table 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 LSS	248.6	260.7	61.6
Snyders	Snyders (Con)	210.2	241.6	52.8
Kellogg	Zoo Animal Cracker Austin	183.1	233.7	96.8
Kraft	Planters (Con)	161.1	218.8	96.0
	Total	4892.1	5357.9	74.4
Zoo Animal Crackers Removal				
Mars	M&M Peanut	309.7	420.3	99.2
Mars	Snickers	301.3	385.1	94.4
Pepsi	Rold Gold (Con)	158.9	342.4	92.0
Snyders	Snyders (Con)	210.3	263.0	67.2
Pepsi	Cheeto LSS	248.6	263.0	66.4
Mars	Twix Caramel	169.1	235.0	99.2
Pepsi	Baked Chips (Con)	169.6	219.7	89.6
	Total	4892.2	5608.6	89.6
Famous Amos Cookie Removal				
Mars	M&M Peanut	309.7	319.5	46.4
Mars	Snickers	301.2	316.6	52.0
Pepsi	Rold Gold (Con)	158.9	285.3	80.0
Pepsi	Cheeto LSS	248.7	260.7	64.8
Snyders	Snyders (Con)	210.1	236.4	52.8
Pepsi	Sun Chip LSS	150.2	225.5	100.0
Pepsi	Ruffles (Con)	206.9	218.3	62.4
	Total	4890.2	5262.4	64.0
M&M Peanut Removal				
Mars	Snickers	300.9	411.8	99.2
Snyders	Snyders (Con)	209.7	279.0	76.8
Pepsi	Rold Gold (Con)	158.9	276.9	80.8
Pepsi	Cheeto LSS	248.6	251.0	47.2
Mars	Twix Caramel	167.9	213.8	90.4
Kellogg	Zoo Animal Cracker Austin	182.6	198.0	65.6
Pepsi	Baked Chips (Con)	169.4	194.7	68.0
	Total	4886.1	5315.5	65.6

Table 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.

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

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

Table 6: Simple Matching Estimator (with and without Assumption 2) (Snickers Removal)

Assumption 1 restricts control machine-weeks to: the same machine, requires that k is available, and that control machine-weeks were not in any treatment.

Assumption 2 requires that removing a product cannot increase total sales during a period, and cannot decrease total sales by more than the expected sales of the removed product)

Trt'd Mach-Weeks shows the number of treated machine-weeks for which there was at least one control machine-week. Avg # Controls Per Trt is the average number of control machine-weeks per treatment machine-week over all treatment machine-weeks.

Δq_k 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.

Manufacturer	Product	Treated Machine Weeks	Δq_k Subst Sales	Δq_j Focal Sales	$\Delta q_k/\lvert\Delta q_j\rvert$ Diversion	w/ Assn 3 Diversion ($m = K$)	w/ Assn 3 Diversion ($m = 300$)	w/ Assn 4 Diversion ($m = 3.3$)
				Snickers Removal				
Nestle	Nonchoc Nestle (Con)	3	9.4	-10.5	89.5	12.4	3.1	0.7
Mars	Twix Caramel	134	289.6	-702.4	41.2	37.9	29.5	15.9
Mars	M&M Peanut	176	375.5	-954.3	39.4	37.0	30.8	18.4
Kellogg	Rice Krispies Treats	17	17.7	-66.5	26.7	13.5	5.0	1.3
Nestle	Butterfinger	61	72.9	-362.8	20.1	17.1	11.2	4.5
Mars	Choc Mars (Con)	11	6.4	-32.7	19.7	6.5	2.0	0.4
Pepsi	Rold Gold (Con)	174	161.4	-900.1	17.9	16.8	13.9	7.5
Hershey	Choc Herhsey (Con)	41	29.8	-179.6	16.6	12.2	6.3	2.0
Mars	M&M Milk Chocolate	97	71.8	-457.4	15.7	13.8	9.8	4.1
Misc	Farleys (Con)	18	14.9	-114.2	13.0	8.3	3.7	1.0
Hershey	Payday	2	1.1	-9.8	10.9	1.4	0.4	0.1
Kraft	Planters (Con)	136	78.0	-759.9	10.3	9.6	7.8	3.8
	Outside Good	180	460.9	-970.2	47.5			23.1
				Zoo Animal Crackers Removal				
Hershey	Payday	2	0.4	-0.4	84.7	0.6	0.1	
Kellogg	Rice Krispies Treats	13	23.5	-37.8	62.2	23.2	7.2	3.0
Misc	Salty United (Con)	6	10.4	-18.9	55.1	12.6	3.4	1.3
Kraft	100 Cal Oreo Thin Crisps	13	14.9	-37.8	39.4	14.7	4.5	1.8
Pepsi	Rold Gold (Con)	132	114.4	-440.8	25.9	22.9	16.2	9.9
Hershey	Choc Herhsey (Con)	30	33.6	-132.6	25.3	17.1	7.9	3.8
Misc	Hostess Pastry	11	14.7	-62.2	23.7	11.8	4.4	1.8
Kraft	100 Cal Chse Nips Crisps	13	8.7	-37.8	23.1	8.6	2.6	1.1
Mars	Milky Way	9	7.0	-30.8	22.6	7.5	2.2	0.9
Mars	Snickers	145	92.4	-483.6	19.1	17.3	13.0	7.6
Mars	M&M Peanut	142	77.7	-469.4	16.6	15.0	11.4	6.5
Mars	Twix Caramel	110	50.2	-339.0	14.8	12.7	8.7	4.6
	Outside Good	145	240.5	-482.9	49.8			22.0
				Famous Amos Cookie Removal				
Nestle	Choc Nestle (Con)	1	0.8	-0.3	300.0	1.2	0.3	
Hershey	Choc Herhsey (Con)	38	48.6	-66.8	72.7	36.9	13.4	7.2
Kraft	100 Cal Oreo Thin Crisps	29	20.7	-43.3	47.9	19.2	6.1	3.1
Pepsi	Sun Chip LSS	139	143.6	-355.7	40.4	34.4	22.7	15.7
Hershey	Payday	2	2.6	6.8	38.9			
Misc	Salty United (Con)	18	9.9	-28.7	34.6	10.7	3.1	1.5
Pepsi	Chs PB Frito Cracker	34	26.9	-83.6	32.1	18.2	7.1	3.7
Kraft	Planters (Con)	121	82.1	-332.6	24.7	20.9	13.7	8.8
Kellogg	Choc SandFamous Amos	57	28.0	-122.0	22.9	15.1	6.8	3.7
Mars	Milky Way	26	13.9	-71.6	19.5	10.3	3.9	1.9
Pepsi	Dorito Blazin Buffalo Ranch LSS	72	38.1	-224.2	17.0	13.3	7.5	4.4
Pepsi	Frito LSS	119	49.9	-313.2	15.9	13.4	8.9	5.3
	Outside Good	156	192.9	-399.1	48.3			21.0

Manufacturer	Product	Treated Machine Weeks	Δq_k Subst Sales	Δq_j Focal Sales	$\Delta q_k/\lvert\Delta q_j\rvert$ Diversion	w/ Assn 3 Diversion ($m = K$)	w/ Assn 3 Diversion ($m = 300$)	w/ Assn 4 Diversion ($m = 3.3$)
				M&M Peanut Removal				
Misc	Hostess Pastry	11	12.5	-38.6	32.5	12.3	4.0	1.8
Mars	Snickers	218	296.6	-1239.3	23.9	22.9	19.9	16.5
Kellogg	Brown Sug Pop-Tarts	10	10.0	-43.5	22.9	9.2	2.9	1.4
Misc	Cliff (Con)	1	0.4	-1.8	22.2	0.6	0.1	0.0
Nestle	Nonchoc Nestle (Con)	1	0.9	-4.6	19.5	1.3	0.3	0.2
Mars	M&M Milk Chocolate	99	73.5	-529.6	13.9	12.5	9.2	6.3
Mars	Twix Caramel	176	110.9	-1014.3	10.9	10.4	8.9	6.8
Kellogg	Rice Krispies Treats	46	22.4	-220.2	10.2	7.9	4.4	2.5
Hershey	Twizzlers	62	33.0	-333.0	9.9	8.3	5.3	3.4
Hershey	Choc Herhsey (Con)	32	15.7	-160.0	9.8	7.0	3.5	1.9
Kellogg	Cherry Pop-Tarts	25	12.5	-160.3	7.8	5.6	2.8	1.6
Mars	Nonchoc Mars (Con)	45	14.6	-201.3	7.3	5.5	3.0	1.7
	Outside Good	218	606.2	-1238.5	48.9			36.3

Table 7: Raw and Bayesian Diversion Ratios.

Treated Machine Weeks shows the number of treated machine-weeks for which there was at least one control machine-week. 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 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 $m = 64 - 66$ as the number of pseudo-observations. The strong prior uses $m = 300$ pseudo-observations.

The final column utilizes Assumptions 1, 2 and 4 with the Dirichlet prior and $m = 3.3$ pseudo-observations.

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

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

Table 8: 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.

Manuf	Product	Mean	2.5^{th} Quantile	25^{th} Quantile	50^{th} Quantile	75^{th} Quantile	97.5^{th} Quantile
			Snickers Removal				
Nestle	Nonchoc Nestle (Con)	0.67	0.31	0.51	0.65	0.81	1.17
Mars	Twix Caramel	15.88	14.28	15.32	15.88	16.45	17.53
Mars	M&M Peanut	18.40	16.79	17.83	18.39	18.95	20.02
Kellogg	Rice Krispies Treats	1.30	0.78	1.09	1.28	1.49	1.95
Nestle	Butterfinger	4.45	3.53	4.10	4.43	4.78	5.48
Mars	Choc Mars (Con)	0.44	0.16	0.31	0.42	0.55	0.85
Pepsi	Rold Gold (Con)	7.54	6.49	7.15	7.53	7.92	8.69
	Outside Good	23.12	21.34	22.50	23.11	23.73	24.91
			Zoo Animal Crackers Removal				
Kellogg	Rice Krispies Treats	2.99	1.93	2.56	2.95	3.36	4.28
Misc	Salty United (Con)	1.25	0.61	0.97	1.21	1.49	2.12
Kraft	100 Cal Oreo Thin Crisps	1.85	1.04	1.51	1.81	2.14	2.88
Pepsi	Rold Gold (Con)	9.89	8.24	9.30	9.88	10.46	11.66
Hershey	Choc Herhsey (Con)	3.81	2.66	3.35	3.77	4.22	5.17
Misc	Hostess Pastry	1.80	1.02	1.47	1.76	2.08	2.79
Kraft	100 Cal Chse Nips Crisps	1.10	0.51	0.83	1.06	1.32	1.91
	Outside Good	21.98	19.64	21.15	21.96	22.78	24.43
			Famous Amos Cookie Removal				
Hershey	Choc Herhsey (Con)	7.18	5.38	6.49	7.14	7.83	9.21
Kraft	100 Cal Oreo Thin Crisps	3.05	1.90	2.59	3.01	3.47	4.47
Pepsi	Sun Chip LSS	15.75	13.53	14.94	15.72	16.52	18.11
Misc	Salty United (Con)	1.47	0.70	1.13	1.42	1.75	2.49
Pepsi	Chs PB Frito Cracker	3.74	2.51	3.25	3.70	4.19	5.21
Kraft	Planters (Con)	8.75	7.04	8.13	8.72	9.35	10.64
Kellogg	Choc SandFamous Amos	3.69	2.47	3.21	3.65	4.12	5.15
	Outside Good	20.95	18.43	20.05	20.94	21.83	23.57
			M&M Peanut Removal				
Misc	Hostess Pastry	1.85	1.00	1.49	1.80	2.17	2.95
Mars	Snickers	16.47	14.83	15.89	16.46	17.04	18.15
Kellogg	Brown Sug Pop-Tarts	1.41	0.69	1.10	1.37	1.68	2.39
Misc	Cliff (Con)	0.00	0.00	0.00	0.00	0.00	0.03
Nestle	Nonchoc Nestle (Con)	0.15	0.00	0.05	0.11	0.21	0.54
Mars	M&M Milk Chocolate	6.26	4.96	5.78	6.25	6.73	7.68
Mars	Twix Caramel	6.76	5.60	6.34	6.74	7.16	7.99
	Outside Good	36.35	34.21	35.61	36.34	37.09	38.47

Table 9: Posterior Distribution of Dirichlet $\alpha_{jk} = \frac{s_j}{1-s_j} + \frac{1.3}{K+1}$, $m_{jk} = 3.3$

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

Proposed Merger	Diversion Direction	Diversion Ratio	Proposed Divestiture	Diversion Ratio Under Divestiture
Mars & Hershey	Snickers to Hershey	2.83	Reese's Peanut Butter Cups	2.83*
	M&M Peanut to Hershey	7.14	Reese's Peanut Butter Cups	5.30
Mars & Kraft	Snickers to Kraft	3.97	Planters Peanuts	0.15
	M&M Peanut to Kraft	4.22	Planters Peanuts	0.62
Mars & Nestle	Snickers to Nestle	5.72	Butterfinger	1.27
	M&M Peanut to Nestle	6.32	Butterfinger	4.52
Mars & Kellogg's	Snickers to Kellogg's	8.56	Famous Amos Cookies†	4.60
	M&M Peanut to Kellogg's	5.67	Famous Amos Cookies†	5.50
	Zoo Animal Crackers to Mars	21.78	Famous Amos Cookies†	21.78
Kellogg's & Kraft	Zoo Animal Crackers to Kraft	5.81	Planters Peanuts	3.38
	Choc Chip Famous Amos to Kraft	11.82	Planters Peanuts	3.07

Table 10: Hypothetical Mergers with Forced Divestitures

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

For upward pricing pressure not to be positive under the assumptions that $p = 0.45$ and $c = .15$, marginal cost reductions must be at least twice as large as the diversion estimates.

A Appendix:

A.1 Diversion Under Parametric Demands

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

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

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

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

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

This means that for any change in p_j from an infinitesimal price increase, up to the choke price of j; the diversion ratio, D_{jk} is constant. This also implies that under linear demand, divergence 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:

$$\begin{aligned} \frac{\Delta \log(Q_k)}{\Delta \log(Q_j)} \approx \underbrace{\frac{\Delta Q_k}{\Delta Q_j}}_{D_{jk}} \cdot \frac{Q_j(\mathbf{p})}{Q_k(\mathbf{p})} &= \frac{\varepsilon_{kj}\Delta\log(p_j)}{\varepsilon_{jj}\Delta\log(p_j)} = \frac{\varepsilon_{kj}}{\varepsilon_{jj}} \\ D_{jk} &\approx \frac{Q_k(\mathbf{p})}{Q_j(\mathbf{p})} \cdot \frac{\varepsilon_{kj}}{\varepsilon_{jj}} \end{aligned} \tag{A.15}$$

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. This implies that the diversion ratio does not depend on the magnitude of the price increase. Here we consider two price increases, an infinitesimal one and an increase to the choke price $p_j \to \infty$.

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

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

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

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

For a price change to the choke price,

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

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

We can also show that the bias expression for the diversion ratio is equal to zero with logit demand (i.e., that $D_{jk} = \frac{\partial^2 q_k}{\partial p_j^2} / \frac{\partial^2 q_j}{\partial p_j^2}$):

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

Random Coefficients Logit Demand

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

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

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

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

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

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

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 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 doesn'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):

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

In both cases we observe substantially less heterogeneity in β_{it}^{price} and we also observe that MTE_p, MTE_q, ATE are more similar to one another in Table 11.

	med$(y-x)$	mean$(y-x)$	med$(\|y-x\|)$	mean$(\|y-x\|)$	std$(\|y-x\|)$
	Nevo (2000) Estimates				
	All Products				
MTE_q	0.62	1.76	2.85	4.57	5.06
ATE	1.05	1.91	3.18	4.97	5.42
Logit	-29.15	-29.09	33.98	40.05	31.60
	Best Substitutes				
MTE_q	0.75	1.56	2.15	3.50	3.88
ATE	1.39	2.45	2.51	4.16	5.00
Logit	-31.83	-35.01	32.72	38.40	29.13
	Outside Good				
MTE_q	-2.32	-2.39	2.67	3.05	2.23
ATE	-2.90	-3.24	3.09	3.76	3.05
Logit	32.52	40.49	32.52	41.02	30.67
	Restricted Estimates $\pi_{inc^2,price} = 0$				
	All Products				
MTE_q	1.37	3.22	6.84	10.51	10.93
ATE	2.02	3.11	7.34	11.12	11.39
Logit	-33.48	-19.13	50.80	56.00	36.46
	Best Substitutes				
MTE_q	1.68	3.93	4.49	6.86	7.36
ATE	2.52	5.26	4.77	7.78	9.02
Logit	-41.56	-40.50	43.23	47.47	29.60
	Outside Good				
MTE_q	-4.51	-5.44	4.55	5.72	4.59
ATE	-5.11	-6.45	5.14	6.70	5.73
Logit	30.46	35.38	30.56	37.05	27.04

Table 11: Alternative Specifications for Nevo (2000).

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

Table 12: Simulation comparing ATE and MTE for Random Coefficients Logit

A.3 Discrepancy Between Average and Marginal Treatment Effects

We can perform a Monte Carlo study to analyze the extent to which the average treatment effect deviates from the marginal treatment effect. We generate data by simulating from a random coefficients logit model with a single random coefficient on price. Our simulations follow the procedure in Armstrong (2013), Judd and Skrainka (2011) and Conlon (2016) where prices are endogenously solved for via a Bertrand-Nash game given the other utility parameters, rather than directly drawn from some distribution.

We generate the data in the following manner: $u_{it} = \beta_0 + x_j\beta_1 - \alpha_i p_j + \xi_j + \varepsilon_{ij}$ and $mc_j = \gamma_0 + \gamma_1 x_j + \gamma_2 z_j + \eta_j$ where $x_j, z_j \sim N(0,1)$, with $\xi_j = \rho\omega_{j1} + (1-\rho)\omega_{j2} - 1$ and $\eta_j = \rho\omega_{j1} + (1-\rho)\omega_{j3} - 1$ and $(\omega_1, \omega_2, \omega_3) \sim^{i.i.d.} U[0,1]$. Following Armstrong (2013) and Conlon (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 in a Bertrand-Nash equilibrium.

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

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

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

Table 13: Monte Carlo Simulations

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

Table 14: Monte Carlo Simulations

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

Table 15: Monte Carlo Simulations

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

Table 16: Monte Carlo Simulations

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).[49] 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 1.3 pseudo observations from a uniform prior $\frac{1}{K+1}$ to each substitute. This gives a Dirichlet parameter of $\alpha_k = \frac{s_k}{1-s_j} + \frac{1.3}{K+1}$. We then choose α_0 so that outside good share $\mu_0 = 0.25$ for the prior distribution. This results in $m = 3.05$ 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).[50]

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 $\alpha_k = \frac{s_k}{1-s_j} + \frac{1.3}{K+1}$ with $m = 3.05$ pseudo-observations. The discrepancies between these priors are reported in Table 17. We obtain nearly identical results (differences less than 0.03 *percentage points*) when

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

[50] 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 $\mu_{jk} = \frac{1}{K+1}$.

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

Table 17: Maximum Absolute Deviation (percentage points) between Dirichlet parametrized by (adjusted) IIA logit shares ($\alpha_k = \frac{s_k}{1-s_j} + \frac{1.3}{K+1}$, $\mu_0 = 0.25$, $m = 3.05$) and alternatives.

compared to the Dirichlet with a the 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 stronger prior rather than the share of the outside good.[51]

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

A.5 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]);
    }
}
```