Empirical Properties of Diversion Ratios

Christopher T. Conlon[y]
Julie Holland Mortimer[z]

January 2, 2019

Abstract

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

An earlier version of this paper was circulated under the title *An Experimental Approach to Merger Analysis.* We thank Mark Stein, Bill Hannon, and the drivers at Mark Vend Company for implementing the experiment used in this paper, providing data, and generally educating us about the vending industry. We thank Dennis Carlton, Joe Farrell, Ken Hendricks, Fiona Scott Morton, Aviv Nevo, Dan O'Brien, Bill Rogerson, Ralph Winter, and seminar participants at Columbia University, NYU Stern, the University of Western Ontario, the FTC Microeconomics Conference, `I.O. Fest' at UC Berkeley, the Dartmouth Winter IO Conference, the NBER Summer Institute, and the antitrust conference at the University of British Columbia for comments. Andrew Copland, Bogdan Genchev, Tom Gole, and Joonwoo Park provided exceptional research assistance. Financial support for this research was generously provided through NSF grant SES-0617896. Any remaining errors are our own.

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

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

1 Introduction

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

Not surprisingly, diversion ratios are a central calculation of interest to antitrust authorities for analyzing horizontal mergers. The current U.S. merger guidelines, released in 2010, place greater weight on diversion ratios relative to concentration measures more commonly used to understand competition in settings with homogeneous goods (e.g., the Her ndahl-Hirschman Index (HHI)).[1]

survey data or in a firm's course of business. A 2017 commentary on retail mergers released by the UK Competition and Markets Authority (CMA) describes their use of diversion ratios for screening and analyzing mergers, saying:

> Diversion ratios can be calculated in a number of different ways, depending on the information available in a particular case. In retail mergers, the CMA has most often used the results of consumer surveys to calculate diversion ratios. The diversion ratio attempts to capture what customers would do in response to an increase in prices. However, it can be difficult to survey a sufficiently large number of customers who would switch in response to a price rise to estimate a robust diversion ratio. Therefore, the CMA asks customers what they would do in response to the closure of a store (or stores).[3]

In this paper, we analyze different ways of estimating diversion ratios and characterize their empirical properties.

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

When policy-makers are interested in measuring the effect of treating only those consumers who substitute away from j after a very small price increase, they are implicitly evaluating a marginal treatment effect (MTE) at pre-merger prices.[4] A challenge of directly implementing such an experiment is that treating a small number of the most price-sensitive individuals may lack statistical power. An alternative is to treat all individuals who would have purchased j at pre-existing prices, and thus estimate an average treatment effect (ATE). This can be accomplished by surveying consumers about their second-choice products, or by exogenously removing product j from the choice set. When the diversion ratio is constant, the ATE coincides with the evaluation of the MTE at pre-merger prices. However, we show that constant diversion is a feature of only the linear demand model and a `plain vanilla' logit model. Other commonly-used models of demand, such as random-coe cients logit or log-linear models, do not feature constant diversion, and the ATE may diverge from the MTE evaluated at pre-merger prices.

A related question for the antitrust authority is whether one can reliably estimate diversion ratios using data from only the merging entities. To consider this question, it's useful to consider two concepts: an aggregate diversion ratio, which Katz and Shapiro (2003) de ne as the \percentage of the total sales lost by a product when its price rises that are captured by all of the other products in the candidate market," and a diversion matrix, which we de ne as a matrix whose o -diagonal elements report diversion between each pair of products that could potentially be considered for inclusion in a market, and whose diagonal elements report diversion to the outside good.[5] Discrete-choice models of demand imply a \summing up" constraint so that each row of the diversion matrix (i.e., aggregate diversion plus diversion the best substitute, and overstates diversion to the outside good compared to both the ATE and MTE measures.

In a third application, we construct an empirical estimator for the ATE measure of the diversion ratio by exogenously removing products from a set of vending machines in a large-scale experiment and tracking subsequent substitution patterns. The experimental setting precludes us from estimating diversion that would be relevant to a small price change because we are not able to exogenously change prices, but it does not require any parametric restrictions or any restrictions on aggregate diversion. In order to control for unobserved demand shocks, we select valid controls and impose a simple requirement that product removals cannot increase total sales, nor decrease total sales by more than the sales of the product removed.

Having matched to these control observations, we consider, in turn, two additional assumptions about economic primitives and examine how they help to estimate experimental measures of the diversion ratio. The rst assumption is that diversion to any single product is between 0 and 100 percent. We incorporate this assumption through a non-parametric Bayesian shrinkage estimator. We nd that this improves our estimates of diversion, although our estimates are sensitive to the strength of the prior. Next, we impose the assumption that aggregate diversion plus diversion to the outside good sums to one. Our Bayesian shrinkage estimator incorporates this assumption by nesting the parametric structural estimates of diversion and the (quasi)-experimental measures in a single framework. With the \summing up" constraint, even a very weak prior yields precise estimates of diversion ratios.

Our results highlight two important points: (1) Observing data from all products within the market, rather than only products involved in a merger, is important when estimating diversion ratios; and (2) in discrete-choice demand systems, the \summing up" constraint may play a more important role for identi cation than the parametric distribution of error terms. Our applications also illustrate the fact that di erent measures of diversion may be relevant and/or available to policy-makers in di erent 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.[7] Other cases have centered around the potential for product discontinuations, such as in hospital and airline networks, and in several business-to-business markets.[8]

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

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

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

1.1 Related Literature

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

By exploring the assumptions required for a credible (quasi)-experimental method of measuring diversion, we connect directly to the theoretical literature discussing the use and measurement of the diversion ratio.[9] Farrell and Shapiro (2010) suggest that rms themselves may track diversion in their `normal course of T -1717.9 .r6s8(that)-227(ynt)-22727(ersion)-435t ofr

sets the price of product j to maximize pro ts: $_j$ = (p_j c_j)

This matrix is useful to make three conceptual points: (a) if all products are substitutes (rather than complements) and consumers make discrete choices, then each row of the matrix must sum to one: $D(p) \quad \mathbf{1}_J = \mathbf{1}_J$; (b) the sum of the o -diagonal elements along a row j is known as aggregate diversiorfor product j;[13]

effect with a binary treatment (i.e., not purchasing product j) and a binary outcome (i.e., purchasing product k or not). We denote this as:

$$D_{jk}(p_j; p^0_{-j}) = -\frac{\Delta q_k}{\Delta q_j} = -\frac{q_k(p_j^0 + \Delta p_j; p^0_{-j}) - q_k(p_j^0; p^0_{-j})}{q_j(p_j^0 + \Delta p_j; p^0_{-j}) - q_j(p_j^0; p^0_{-j})} = -\frac{\int_{p_j^0}^{p_j^0+\Delta p_j} \frac{\partial q_k(p_j; p^0_{-j})}{\partial p_j} dp_j}{\int_{p_j^0}^{p_j^0+\Delta p_j} \frac{\partial q_j(p_j; p^0_{-j})}{\partial p_j} dp_j} \quad (2)$$

The treated group corresponds to individuals who would have purchased product j at price p_j^0 but do not purchase j at price $(p_j^0 + \Delta p_j)$. The lower an individual's reservation price for j, the more likely an individual is to receive the treatment. Thus, Δp_j functions as an `instrument' because it monotonically increases the probability of treatment.

By focusing on the numerator in equation (2), we can re-write the diversion ratio using the marginal treatment effects (MTE) framework of Heckman and Vytlacil (2005), in which $D_{jk}(p_j; p^0_{-j})$ is a marginal treatment effect that depends on p_j.[15]

$$\widehat{D}^{LATE}_{jk}(\Delta p_j) = \frac{1}{\Delta q_j} \int_{p_j^0}^{p_j^0+\Delta p_j} \underbrace{-\frac{\partial q_k(p_j; p^0_{-j})}{\partial q_j}}_{D_{jk}(p_j; p^0_{-j})} \frac{\partial q_j(p_j; p^0_{-j})}{\partial p_j} \partial p_j \quad (3)$$

$$\widehat{D}^{ATE}_{jk} = \frac{1}{\Delta q_j} \int_{p_j^0}^{\overline{p}_j} D_{jk}(p_j; p^0_{-j}) \frac{\partial q_j(p_j; p^0_{-j})}{\partial p_j} \partial p_j = -\frac{q_k(\overline{p}_j; p^0_{-j}) - q_k(p_j^0; p^0_{-j})}{q_j(\overline{p}_j; p^0_{-j}) - q_j(p_j^0; p^0_{-j})} \quad (4)$$

As we vary p_j, we measure the weighted average of diversion ratios where the weights $w(p_j) = \frac{1}{\Delta q_j} \frac{\partial q_j(p_j; p^0_{-j})}{\partial p_j}$ correspond to the lost sales of j at a particular p_j as a fraction of all lost sales. This directly corresponds to Heckman and Vytlacil (2005)'s expression for the local average treatment effect (LATE); we average the diversion ratio over the set of consumers of product j who are most price sensitive. The LATE estimator varies with the size of the price increase because the set of treated individuals varies. Equation (3) confirms that the LATE estimate concentrates more weight near p^0 when demand is more elastic, or when demand becomes increasingly inelastic for larger Δp_j. In equation (4) the average treatment effect (ATE) is the expression for the LATE in which all individuals are treated. This corresponds to an increase in p_j

for Δp_j our LATE estimate may differ from the MTE evaluated at $D_{jk}(p^0)$.

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

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

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

This allows us to compute an expression for the difference between a LATE estimate $\widehat{D}_{jk}^{LATE}(\Delta p_j)$ and the MTE evaluated at $D_{jk}(p^0)$. We refer to this as the `bias' of the LATE estimate:[17]

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

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

The expression in equation (6) shows that the bias (i.e., the difference between the two estimates of diversion) depends on two things: the magnitude of the price increase Δp_j, and the cur 0 G Buehr Tf 4.382 1.794 Td [(,)-303(and)]e [(,)-303(and)]e7(di -303(and)]TJ -12]03(and(et)and

Equation (6) also provides insight into the economic implications of assuming a constant treatment effect (i.e., assuming that $D_{jk}(p_j; p_{-j}) = D_{jk}$). Constant diversion requires that the bias calculation in equation (6) is equal to zero. Two functional forms for demand exhibit constant diversion: linear demand, for which $\frac{\partial^2 q_k}{\partial p_j^2} = 0, \forall j; k$; and the IIA logit model, for which $D_{jk} = \frac{\partial^2 q_k}{\partial p_j^2} / \frac{\partial^2 q_j}{\partial p_j^2}$. Implicitly when we assume that the diversion ratio does not vary with price, we assume that the true demand system is well approximated by either a linear demand curve or the IIA logit model.We derive these relationships, as well as expressions for diversion under other demand models in Appendix A.1, and show that random-coefficients logit demand, and constant elasticity demands (including log-linear demand) do not generally exhibit constant diversion.

To summarize, we can expect a LATE or ATE measure of diversion to be similar to the MTE evaluated at p^0 when the bias in (6) is small. This happens when: (a) the changes. Although diversion to the three alternatives at any given price point is the same as the case of inelastic demand, the ATE measure of diversion is now more heavily weighted towards consumers that leave at small price changes (72% to Honda Civic, 10% to Tesla, and 18% to the outside good).

2.3 Utilizing Second-Choice Data to Measure Diversion

Often researchers have access to second-choice data. For example, Berry, Levinsohn, and Pakes (2004) observe not only market shares of cars, but also survey answers to the question: \If you did not purchase this vehicle, which vehicle would you purchase?" Consumer surveys provide a stated-preference method of recovering second-choice data. One may also construct second-choice data through a revealed-preference mechanism by experimentally removing product j from a consumer's choice set for a period of time.[19] One can view such an exogenous product removal as being equivalent to an increase in price to the choke price $\bar{p}_j$, where $q_j(\bar{p}_j; p_{-j}) = 0$. Thus, an exogenous product removal measures the ATE, treating all of the pre-merger consumers of good j and minimizing the variance expression in (7).

Notice the relationship between the ATE measure of diversion $\widehat{D}^{ATE}_{jk}$ 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:

$$\widehat{D}^{ATE}_{jk} = \frac{q_k(\bar{p}_j; p^0_{-j}) - q_k(p^0_j; p^0_{-j})}{q_j(\bar{p}_j; p^0_{-j}) - q_j(p^0}$$

mimic the cereal industry from ??.[20] These applications allows us to measure diversion in two ways: first, as the ratio of the estimated derivatives of demand evaluated at pre-merger prices (a MTE evaluated at p_0), and second, as the response to a simulated removal of a product (an ATE). The discrete-choice nature of the demand system imposes a `summing-up constraint' (i.e., that aggregate diversion plus diversion to the outside good sums to one). The exercise is meant to demonstrate that the ATE and MTE measures of diversion may not always coincide under commonly-used parametric forms of demand, and to allow us to analyze characteristics of demand that may cause ATE and MTE measures to diverge.

3.1 Nevo (2000)

The data from Nevo (2000) cover sales of ready-to-eat (RTE) cereal in $T = 94$ markets with $J = 24$ products per market.[21] Nevo (2000) allows for a $I = 20$ point distribution of heterogeneity for each market, product fixed effects d_j, unobserved heterogeneity in the form of a multivariate normally distributed ν_i with variance Σ, and observable demographic heterogeneity in the form of Π interacted with a vector of demographics d_{it}.

$$u_{ijt} = d_j + x_{jt}\underbrace{(\bar{\beta} + \Sigma\nu_i + \Pi d_{it})}_{\alpha_{it}} + \xi_{jt} + \varepsilon_{ijt}$$

We estimate parameters following the MPEC approach of Dube, Fox, and Su (2012).[22] The estimated coefficient on price is distributed as follows:[23]

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

We denote a measure of diversion evaluated for an infinitesimally small price change as a MTE. We refer to a `second choice' estimate of diversion as an ATE. For comparison, we also evaluate a Logit model, under which diversion is assumed to be constant. In Appendix A.1 we derive these measures for commonly-used parametric forms. For the parametric forms

[20] These data are posted online by the author, and are not the actual data used in ??, which are proprietary.

[21] The data that Nevo (2000) is able to make available for replication exclude product and market names, so we cannot reference specific product names or markets in our analysis.

[22] 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.

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

3.2 Berry, Levinsohn, and Pakes (1999)

The data from Berry, Levinsohn, and Pakes (1999) cover sales of automobiles in the U.S. and consist of $T = 21$ markets (each market is a year) with up to $J = 150$ products per market and a total of 2,217 product-market pairs. The model allows for random tastes for: vehicle size, miles-per-dollar, air conditioning, horsepower per unit of weight, and a constant. It also allows for a coe cient on price that depends on income. The model includes simultaneous estimation of both supply and demand.

Tables 3 and 4 repeat the exercises in tables 1 and 2. Relative to the market for RTE cereal, the market for automobiles features much greater price variation, and potentially an opportunity for greater variation between the MTE and ATE measures of diversion. In table 3, we see that overall levels of diversion to the best substitute are lower for the auto application (about 6% on average) than for the RTE cereal application (15% on average), and diversion to the best substitute for the logit model is very low, at less than one percent (compared to 10% on average for RTE cereal). Table 4 shows that the percentage di erence between the MTE and ATE measures of diversion are indeed much larger than those for RTE cereal. For the best substitute, the absolute di erence between the MTE and ATE measures of diversion is 12:6% on average (with a median absolute di erence of 11:5%), whereas averaging over all inside products gives an absolute di erence between the two measures of 40:3% on average (with a median absolute di erence of 22:5%). Diversion to the outside good di ers across the two measures by 25% on average. Diversion under the logit model is wildly di erent from both the MTE and ATE measures; on average, it di ers by 240% for the best substitute and 177% across all products.

In a practical sense, the most important di erence between the RTE cereal and automobile applications is likely the amount of price variation in the market. The fact that the qualities, costs, and prices of autos vary so much more than qualities, costs, and prices of breakfast cereals provides an opportunity to observe larger di erences in the diversion between marginal and inframarginal consumers of autos.

The ATE measure may either overstate or understate diversion to other products on average compared to the MTE measure. If the marginal consumer tends to become more (less) inelastic as the price increases, then the ATE will overstate (understate) substitution.[25]

[25] The elasticity of the marginal consumer will depend on the curvature of demand. For a plain vanilla logit model, the logit error term implies that the elasticity of the marginal consumer increases with price. However, this need not hold for other models of demand. For example, a random coe cient logit model has an in ection point when market share exceeds 0.5. At the market level, $s_j < 0{:}5$ for all j except for the outside good. Empirically, the outside good share may be less than 0.5 in some markets, but greater than

Reducing an estimator's ability to accommodate heterogeneity in consumer preferences produces MTE and ATE measures that are closer together. We demonstrate this e ect with Monte Carlo simulations of commonly-used parametric demand models in Appendix A.3.

4 Empirical Application to Vending

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

4.1 Description of Data and Industry

Globally, the snack foods industry is a $300 billion market annually, composed of a number of large, well-known rms 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 rm 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.[26] We report HHI's at both the category level and for all vending products in table 5 from the midwest region of the U.S. If the relevant market is de ned at the category level, all categories are considered highly concentrated, with HHIs in the range of roughly 4500-6300. If the relevant market is de ned as all products sold in a snack-food vending machine, the HHI is below the critical threshold of 2500. Any evaluation of a merger in this industry would hinge on the closeness of competition.

Over the last 25 years, the industry has been characterized by a large amount of merger and acquisition activity, both on the level of individual brands and entire rms. For example, the Famous Amoscookie brand was owned by at least seven rms 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 Crackershave 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).[27]

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 con guration.

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.[28] 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.[29]

[27]Snack foods have an important historic role in market de nition. 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 signi cant markets, as Drake's had only a 2% market share 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.]

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

[29]

observations by week, because di erent visits occur on di erent days of the week.[34] The cost of implementing the experiment consisted primarily of drivers' time.[35]

5 A Nonparametric Estimator for Diversion

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

Using four simple assumptions motivated by economic theory, we develop an estimator for the average treatment effect version of the diversion ratio which deals with these challenges. The first two assumptions restrict the set of machine-weeks that can act as a control for a particular machine-treated week as in a matching estimator. The second two assumptions affect the way in which estimates of $\widehat{\Delta q_j}, \widehat{\Delta q_k}$ are used to construct $\overline{D}_{jk}$ and employ the principle of Bayesian Shrinkage. All four assumptions are implications of the economic restriction that consumers make discrete choices among substitutes.[38] Using the following four assumptions we demonstrate how we estimate $\overline{D}_{jk}$ from our experimental data.

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

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

We adjust our calculation of the expectation to address the volatility of demand.[40] To be explicit, one can introduce a covariate (demand shock):

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

The treated and control periods have different distributions of covariates (demand shocks) because $f(\ |W = 1) \neq f(\ |W = 0)$. The typical solution involves matching or balancing where one re-weights observations in the control period using measure $g(\ )$ so that $f(\ |W = 1) = g(\ |W = 0)$ and then calculates the expectation $E_g[q_k|Z = 0]$ with respect to measure g.[41] For each treated week t, one can construct a set of matched control weeks within a neighborhood $S(\ _t)$, where $S(\ _t)$ is the set of control weeks that correspond to treated week t, and $\ _t$ is an unobserved demand shock. Having chosen $S(\ _t)$, the change in sales for the chosen control weeks is given as:

$$q_{k,t}(\ _t) = q_{k,t}(\ _t) \quad \frac{1}{|\#s \in S(\ _t)|} \sum_{s \in S(\ _t)} q_{k,s} \quad \text{with} \quad d q_k = \sum_t \quad q_{k,t}(\ _t) \qquad (9)$$

Our first two assumptions tell us how to choose $S(\ _t)$. Assumption 1 is straightforward: it controls for unobserved machine-level heterogeneity by restricting potential controls to different (untreated) weeks at the same machine. If $\ _t$ were observed, one could employ conventional matching estimators (such as k-nearest neighbor or local-linear regression (Abadie and Imbens 2006)). However, $\ _t$ is unobserved, so we rely on Assumption 2 (removing a product cannot increase total sales, and cannot reduce sales by more than the sales of the product removed) instead.

We implement Assumption 2 as follows. We let Q_t denote the sales of all products during the treated machine-week, and Q_s

us to understate the diversion ratio. We propose a modification of (10) that is unbiased. We replace q_{js} with $\hat{q}_{js} = E[q_{js} | Q_s; W = 0]$. An easy way to obtain the expectation is to run an OLS regression of q_{js} on Q_s using data only from untreated machine-weeks that satisfy Assumption 1:

$$S_t \quad \{s : Q^0_s \quad Q^1_t \in [0; \hat{b}_0 + \hat{b}_1 Q^0_s]\} \tag{11}$$

Thus (11) defines the set of control periods S_t that correspond to treatment period t under Assumptions 1 and 2.[42] Plugging this into equation (9) gives estimates of $\widehat{\Delta q_k}$ and $\widehat{\Delta q_j}$.[43]

5.2 Bayesian Shrinkage Assumptions

Our Assumptions 3 and 4 place restrictions on how we calculate the diversion ratio given our estimates of $\widehat{\Delta q_k}$ and $\widehat{\Delta q_j}$. The idea is that there may be better estimates of $\overline{D}_{jk}$ than the simple ratio $\frac{\widehat{\Delta q_k}}{\widehat{\Delta q_j}}$. For example, we might find large but noisy estimates of diversion to a substitute product based on only a few observations and a better estimate might adjust for that uncertainty.[44]

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

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

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

We implement Assumption 3 by placing a prior on D_{jk} that restricts all of the mass to the unit interval $D_{jk} | \alpha_{jk}; m_{jk} \sim Beta(\alpha_{jk}; m_{jk})$. Assumption 4 goes further and restricts the vec-

42

tor D_j to the unit simplex, which we implement with the prior D_j Dirichlet ($_{j0}$; $_{j1}$; : : : ; $_{jK}$; m_{jk}). This has the e ect of using information about D_{jk} to inform our estimates for D_{jk^0}.

There are two ways to parametrize the Beta (and Dirichlet) distributions. In the traditional Beta($_1$; $_2$) formulation $_1$ denotes the number of prior successes and $_2$ denotes the number of prior failures (observed before any experimental observations). Under an alternative formulation, Beta(; m): $= \frac{_1}{_1 + _2}$ denotes the prior mean and m denotes the number of \pseduo-observations"m = $_1 + _2$. We work with the latter formulation for both the Beta and Dirichlet distributions.[45] This formula makes it easy to express the posterior mean (under Assumption 3) as a shrinkage estimator that combines our prior information with our experimental data:

$$\hat{D}_{jk} = \ _{jk} + (1 \ ) \frac{q_k}{q_j}; \quad = \frac{m_{jk}}{m_{jk} + q_j} \tag{13}$$

The weight put on our prior mean is denoted by , and directly depends on how many \pseudo-observations" we observe from our prior before observing experimental outcomes. One reason this estimator is referred to as a \shrinkage" estimator, is because as q_j becomes smaller (and our experimental outcomes are less informative), $\hat{D}_{jk}$ shrinks towards $_{jk}$ (from either direction). Thus, when our product removals provide lots of information about diversion from j to k we rely on the experimental outcomes, but when our experimental variation is less informative, we rely more on our prior information.[46] This has the desirable property of taking extreme but imprecisely estimated parameters and pushing them towards the prior mean.

Our remaining challenge is how to specify the prior ($_{jk}$; m_{jk}). Ideally, the location of the prior would be largely irrelevant while the prior strength m would be as small as possible.[47] An uniform or uninformative prior might be to let $_{jk} = \frac{1}{K+1}$ where K is the

6.2 Merger Evaluation

respectively). Products outside of the merging firms with high diversion are not reported in table 12, but are listed in table 8. These include Nestle's Butterfinger, Kraft's Planters Peanuts, and PepsiCo's Rold Gold and Sun Chips snack brands.[58] Recalling the estimates of diversion from table 8, the most important substitutes for Peanut M&Ms are already owned by Mars, as Snickers, Plain M&M's, and Twix comprise diversion of roughly 30%.

Table 13 reports diversion from Kellogg's top two products to Mars' brands. Once again, the `adding up' constraint of Assumption 4 reduces overall diversion to Mars' brands from 52% to 22% in the case of Zoo Animal Crackers, and corrects a negative estimate of overall diversion in the case of Famous Amos cookies. Estimates of diversion without Assumption 4 identify Milky Way as having high diversion for both of Kellogg's products (23% and 19% diversion under no prior for Animal Crackers and Famous Amos, respectively); applying Assumption 4 reduces these estimates to less than 2%. The degree of diversion from Zoo Animal Crackers to Snickers, Plain and Peanut M&Ms, and Twix Caramel is considerable even with Assumption 4 (a total of about 22%), so one might worry about the potential for a price increase on Zoo Animal Crackers. Products outside of the merging firms that have high diversion from Zoo Animal Crackers and Famous Amos cookies are similar to those for Mars' flagship products: PepsiCo's Rold Gold pretzels (for Animal Crackers), and PepsiCo's economic primitives such as the curvature of demand, whereby the average diversion ratio from second-choice data (ATE) is a good approximation for the MTE.

We explore the empirical properties of diversion ratios in three applications. In the first two, we estimate discrete-choice models of demand using data from Nevo (2000) and Berry, Levinsohn, and Pakes (1999). In the third, we analyze a randomized field experiment, in which we exogenously remove products from consumers' choice sets and measure the ATE directly.

We develop a simple method to recover the diversion ratio from data, which enables us to combine both experimental and quasi-experimental measures with structural estimates as prior information. A non-parametric Bayes shrinkage approach enables us to use prior information (or potentially structural estimates) when experimental measures are not available, or when they are imprecisely measured, and to rely on experimental measures when they are readily available. This facilitates the combination of both first- and second-choice consumer data. We show that these approaches are complements rather than substitutes, and we find benefits from measuring diversion not only between products involved in a proposed merger, but also from merging products to non-merging products.

Our hope is that this makes a well-developed set of quasi-experimental and treatment effects tools available and better understood to both researchers in industrial organization and antitrust practitioners. While the diversion ratio can be estimated in different ways, researchers should think carefully about (1) which treatment effect their experiment (or quasi-experiment) is actually identifying; and (2) what the identifying assumptions required for estimating a diversion ratio implicitly assume about the structure of demand.

Competition and Markets Authority (2017): \Retail Mergers Commentary," April 10, https://www.gov.uk/government/publications/retail-mergers-commentary-cma62.

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): \E ects of Product Availability: Experimental Evidence," NBER Working Paper 16506.

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

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

Dub

Heckman, J. J., and E. Vytlacil (2005): \Structural Equations, Treatment E ects, and Econometric Policy Evaluation," Econometrica, 73(3), 669{738.

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

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

Josephs, L. (2018): \`We wanted to go rst.' Here's what's di erent in the decade since Deltc{Lp]TJ5(merger)-326(with)]TJNoourt-326enend(ted)2386(the)238airlineer

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.

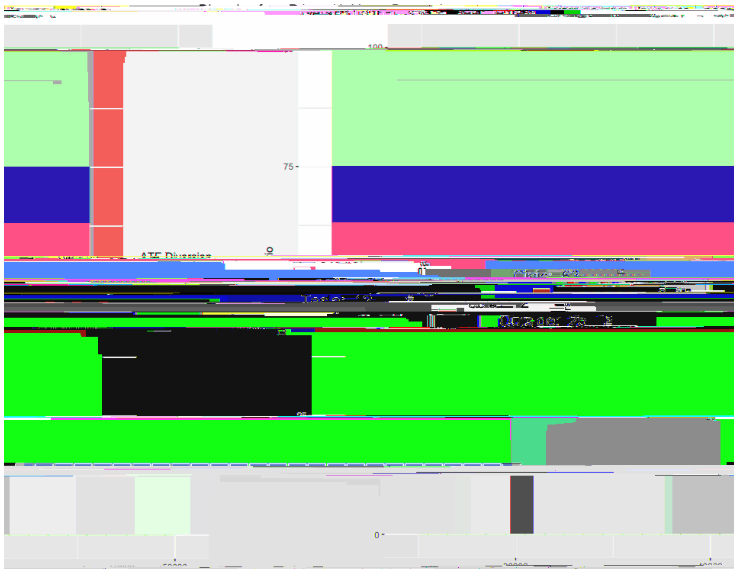
Stock, J. (2010): "The Other Transformation in Econometric Practice: Robust Tools for Inference," The Journal of Economic Perspectives, 24(2), 83–94.

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

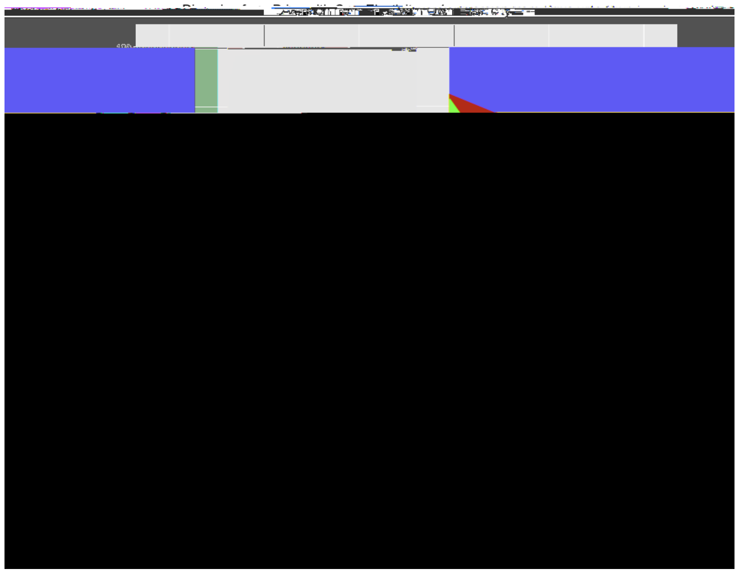
Werden, G. J. (1996): "A Robust Test for Consumer Welfare Enhancing Mergers among Sellers of Differentiated Products," Journal of Industrial Economics, 44(4), 409–13.

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

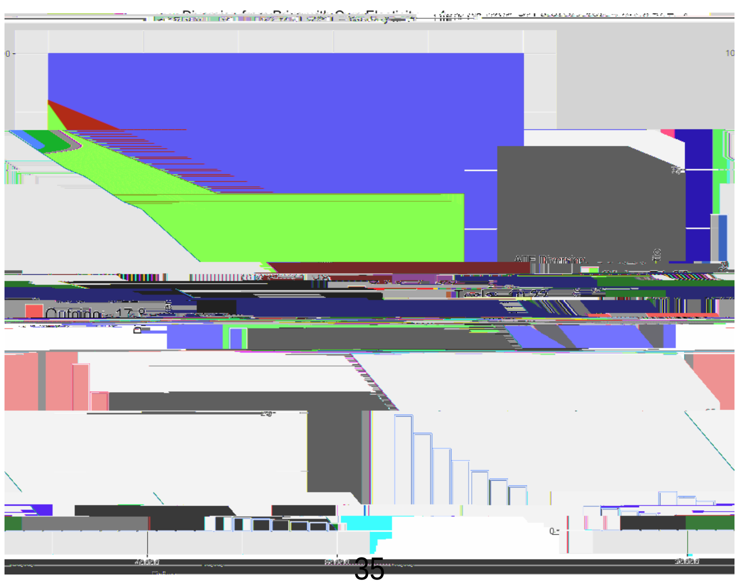
Willig, R. (2011): "Unilateral Competitive Effects of Mergers: Upward Pricing Pressure, Product Quality, and Other Extensions," Review of Industrial Organization, 39(1-2), 19–38.



(a) Linear Demand



(b) Inelastic CES Demand



(c) Elastic CES Demand

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

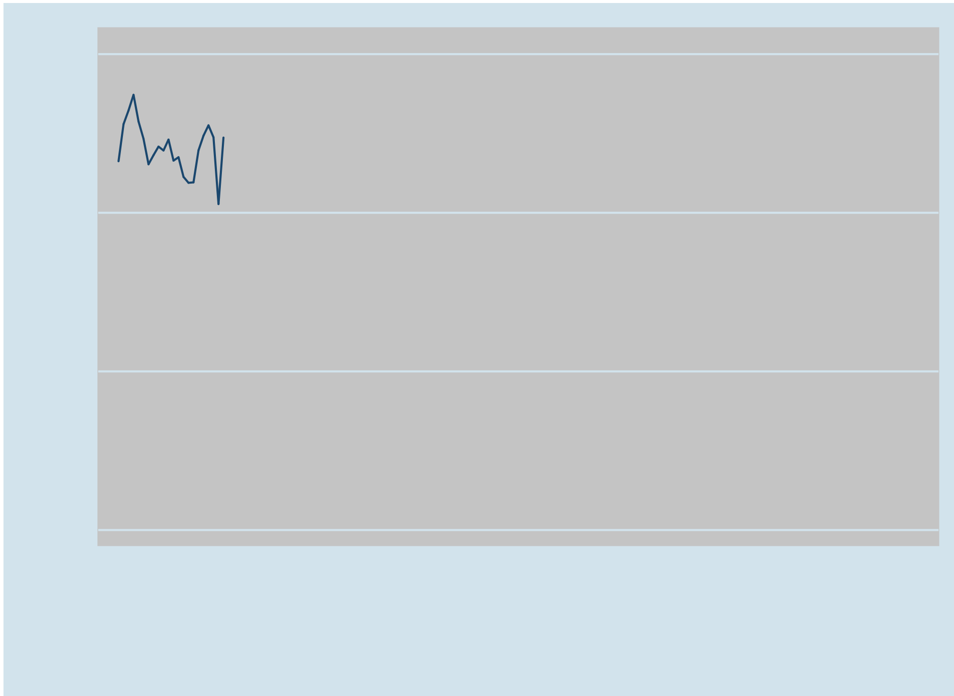


Figure 2: Total Overall Sales and Sales of Snickers and M&M Peanuts by Week

	MTE	ATE	Logit
	Best Substitute		
Med(D_{jk})	5.10	5.04	0.46
Mean(D_{jk})	6.07	6.25	0.53
% Agree with MTE	100.00	96.89	95.62
	Outside Good		
Med(D			

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

Manufacturer	Product	Control Mean	Treatment Mean	Treatment Quantile
		Snickers Removal		
Mars	M&M Peanut	309.8	472.5	100.0
Pepsi	Rold Gold (Con)	158.9	331.9	91.2
Mars	Twix Caramel	169.0	294.1	100.0
Pepsi	Cheeto	248.6	260.7	61.6
Snyders	Snyders (Con)	210.2	241.6	52.8
Kellogg	Zoo Animal Cracker	183.1	233.7	96.8
Kraft	Planters (Con)	161.1	218.8	96.0
	Total	4892.1	5357.9	74.4
		M&M Peanut Removal		
Mars	Snickers	300.9	411.8	99.2
Snyders	Snyders (Con)	209.7	279.0	76.8
Pepsi	Rold Gold (Con)	158.9	276.9	80.8
Pepsi	Cheeto	248.6	251.0	47.2

Mfg	Product	Treated Machine Weeks	q_k Subst Sales	q_j Focal Sales	q_k / j q_j j Div	Assn 3 Diversion (m = K)	Assn 3 Diversion (m = 300)	Assn 4 Diversion (m = 4:15)
				Snickers Removal				
Mars	M&M Peanut	176	375.5	-954.3	39.4	37.0	30.8	18.4
Mars	Twix Caramel	134	289.6	-702.4	41.2	37.9	29.5	15.9
Pepsi	Rold Gold (Con)	174	161.4	-900.1	17.9	16.8	13.9	7.5
Nestle	Butter nger	61	72.9	-362.8	20.1	17.1	11.2	4.5
Mars	M&M Milk Chocolate	97	71.8	-457.4	15.7	13.8	9.8	4.1
Kraft	Planters (Con)	136	78.0	-759.9	10.3	9.6	7.8	3.8
Kellogg	Zoo Animal Cracker	177	65.7	-970.2	6.8	6.5	5.7	2.9
Pepsi	Sun Chip	159	45.3	-866.1	5.2	5.0	4.3	2.1
Hershey	Choc Hershey (Con)	41	29.8	-179.6	16.6	12.2	6.3	2.0
Kellogg	Rice Krispies Treats	17	17.7	-66.5	26.7	13.5	5.0	1.3
Misc	Farleys (Con)	18	14.9	-114.2	13.0	8.3	3.7	1.0
Nestle	Nonchoc Nestle (Con)	3	9.4	-10.5	89.5	12.4	3.1	0.7
Mars	Choc Mars (Con)	11	6.4	-32.7	19.7	6.5	2.0	0.4
Hershey	Payday	2	1.1	-9.8	10.9	1.4	0.4	0.1
Mars	3-Musketeers	2	0.0	0.0				
Misc	BroKan (Con)	3	0.0	0.0				
	Outside Good	180	460.9	-970.2	47.5			23.1
				M&M Peanut Removal				
Mars	Snickers	218	296.6	-1239.3	23.9	22.9	19.9	16.5
Mars	Twix Caramel	176	110.9	-1014.3	10.9	10.4	8.9	6.8
Mars	M&M Milk Chocolate	99	73.5	-529.6	13.9	12.5	9.2	6.3
Nestle	Raisinets	181	71.8	-1001.1	7.2	6.8	5.8	4.4
Kraft	Planters (Con)	190	61.4	-1046.1	5.9	5.6	4.9	3.6
Hershey	Twizzlers	62	33.0	-333.0	9.9	8.3	5.3	3.4
Kellogg	Rice Krispies Treats	46	22.4	-220.2	10.2	7.9	4.4	2.5
Pepsi	Frito	160	37.2	-902.4	4.1	4.0	3.5	2.4
Misc	Hostess Pastry	11	12.5	-38.6	32.5	12.3	4.0	1.8
Kellogg	Brown Sug Pop-Tarts	10	10.0	-43.5	22.9	9.2	2.9	1.4
Nestle	Nonchoc Nestle (Con)	1	0.9	-4.6	19.5	1.3	0.3	0.2
Misc	Cli (Con)	1	0.4	-1.8	22.2	0.6	0.1	0.0
	Outside Good	218	606.2	-1238.5	48.9			36.3

Table 8: Raw and Bayesian Diversion Ratios, Mars' Products

Notes: Treated Machine Weeks shows the number of treated machine-weeks for which there was at least one control machine-week. q_k Subst Sales shows the change in substitute product sales from the control to the treatment period, while q_j Focal Sales shows the analogous change for focal product sales. q_j =j q_j j Diversion is the ratio of the change in substitute product sales to the absolute value of the change in focal product sales. Beta-Binomial (Weak Prior) Diversion and Beta-Binomial (Strong Prior) Diversion are diversion ratios calculated under Assumptions 1, 2 (Substitutes), and 3 (Unit UKi2ovnitPnittBar1-1700(-900.1)-1694(17.9)-3

Mfg	Product	Treated Machine Weeks	q_k Subst Sales	q_j Focal Sales	q_k / j q_j j Div	Assn 3 Diversion (m = K)	Assn 3 Diversion (m = 300)	Assn 4 Diversion (m = 4:15)
Zoo Animal Crackers Removal								
Pepsi								

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

Table 10: Summary Statistics for Diversion Estimates across Products

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

Manuf	Product	Mean	2:5th Quantile	25th Quantile	50th Quantile	75th Quantile	97:5th Quantile
Snickers Removal							
Mars	M&M Peanut	18.40	16.79	17.83	18.39	18.95	20.02
Mars	Twix Caramel	15.88	14.28	15.32	15.88	16.45	17.53
Pepsi	Rold Gold (Con)	7.54	6.49	7.15	7.53	7.92	8.69
Nestle	Butter nger	4.45	3.53	4.10	4.43	4.78	5.48
Kellogg	Rice Krispies Treats	1.30	0.78	1.09	1.28	1.49	1.95
Nestle	Nonchoc Nestle (Con)	0.67	0.6o9				

075 d 0 J 0.398 w 0 0 m 0 24 l S l S Q BT /F60 9.9626 Tf 78.376 569.19 496.6235rs075 d 0 J 0.398 w 0 0 m 0 24 l S l S Q BT /F60 1 72.

	Treated Machine Weeks	No Prior	Assn 3 Diversion (m = K)	Assn 4 Diversion (m = 4:15)
Snickers to Kellogg's Products				
Zoo Animal Cracker	177	6.77	6.53 (5.13; 8.09)	2.92 (2.27; 3.65)
CC Famous Amos	180	4.61	4.46 (3.29; 5.79)	1.99 (1.40; 2.69)
Choc Sandwich FA	69	8.39	7.31 (5.18; 9.77)	1.98 (1.45; 2.59)
Rice Krispies Treats	17	26.68	14.18 (8.66; 20.66)	1.31 (0.79; 1.96)
Cheez-It Original SS	150	0.26	0.35 (0.07; 0.82)	0.10 (0.01; 0.27)
Pop-Tarts*	162	-4.28	0.10 (0.00; 0.38)	0.00 (0.00; 0.02)
Total (to Kellogg's)		42.44	32.93 (26.54; 40.04)	8.30 (7.14; 9.56)
Outside Good	180	47.50	46.19 (43.17; 49.27)	23.19 (21.39; 24.99)
Peanut M&M to Kellogg's Products				
Rice Krispies Treats	46	10.16	7.87 (5.12; 11.12)	2.74 (1.73; 3.95)
CC Famous Amos	215	0.27	0.30 (0.10; 0.66)	0.17 (0.04; 0.40)
Cheez-It Original SS	188	-4.81	0.09 (0.00; 0.31)	0.00 (0.00; 0.03)
Zoo Animal Cracker	218	-2.62	0.10 (0.01; 0.32)	0.00 (0.00; 0.04)
Pop-Tarts*	191	-1.80	0.07 (0.00; 0.27)	0.00 (0.00; 0.03)
Choc Sandwich FA	70	-0.89	0.05 (0.00; 0.34)	0.00 (0.00; 0.03)
Total (to Kellogg's)		0.30	8.46 (5.71; 11.74)	2.92 (1.89; 4.15)
Outside Good	218	48.95	47.85 (45.26; 50.33)	37.62 (35.45; 39.82)

Table 12: Divestitures: Diversion from Mars to Kellogg's products.

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

	Treated Machine Weeks	No Prior	Assn 3 Diversion (m = K)	Assn 4 Diversion (m = 4:15)

A Appendix:

A.1 Diversion Under Parametric Demands

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

Linear Demand

The diversion ratio under linear demand has the property that it does not depend on the magnitude of the price increase. We specify linear demand as:

$$Q_k = {}_k + \sum_j {}_{kj} p_j :$$

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

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

Thus, for any change in p_j from an in nitesimal price increase up to the choke price of j, the diversion ratio D_{jk} is constant. This also implies that under linear demand, diversion is a global property. Any magnitude of price increase evaluated at any initi308(an)2otedan iratiaresvill nesdults0

This holds for small changes in p_j. However for larger changes in p_j we can no longer use the simpli cation that $\log(Q_j) \quad \frac{Q_j}{Q_j}$. So for a large price increase (such as to the choke price p_j ! 1), log-linear demand can exhibit diversion that depends on the magnitude of the price increase.

IIA Logit Demand

The plain logit model exhibits IIA and proportional substitution, which implies that the diversion ratio does not depend on the magnitude of the price increase. We consider two price increases: an in nitesimal one and an increase to the choke price p_j ! 1 . The derivation of the diversion ratio D_{jk} under an IIA logit demand model uses a utility speci cation and choice probabilities given by well-known equations, where a_t denotes the set of products available in market t:

$$u_{ijt} = \underbrace{x_{jt} \quad p_{jt}}_{\forall_{jt}} + "_{ijt}$$

$$S_{jt} =$$

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

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

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

Nested Logit Demand

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

$$\ln s_{jt} \quad \ln s_{0t} = x_{jt} \quad p_{jt} + \ln s_{j|j \in j})(S_j$$

price increase might see substitution from a larger set of consumers. If price sensitivity is correlated with other tastes, then the diversion ratio could differ with the magnitude of the price increase.

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

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

is not identi ed. The easiest choice of a non-price z_{jt} is $_{jt}$, the unobserved product quality term. The role of $_{z}$ is to determine how many individuals receive the treatment as we vary the instrument, but this matters neither in the in nitesimal 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 in nitesimal 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 e ects literature, where di erent instruments trace out di erent MTEs. Thus, the single index of the logit family places an important restriction on the treatment e ects (which may or may not be reasonable).

A.2 Alternative Speci cations for Nevo (2000) Example

Here we repeat the same exercise as in section 3 from the text, but with di erent parameter estimates. In the rst case we use the original published estimates from Nevo (2000) where $_{it}^{price}$ 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: $_{inc^2;price} = 0$. We report those parameter estimates below as well as the estimates in the text from Dube, Fox, and Su (2012):

$$\text{DFS (2012):} \quad {}_{it}^{price} \quad N(\; 62{:}73 + 588{:}21 \; \text{income}_{it} \quad 30{:}19 \; \text{income}_{it}^2 \; + 11{:}06 \; I[\text{child}]_{it} ; \; = 3{:}31)$$
$$\text{Nevo(2000):} \quad {}_{it}^{price} \quad N(\; 32{:}43 + 16{:}60 \; \text{income}_{it} \quad 0{:}66 \; \text{income}_{it}^2 \; + 11{:}63 \; I[\text{child}]_{it} ; \; = 1{:}85)$$
$$\text{Restricted:} \quad {}_{it}^{price} \quad N(\; 34{:}09 + 8{:}53 \; \text{income}_{it} \; + 18{:}16 \; I[\text{child}]_{it} ; \; = 1{:}04)$$

We report both cases in table 14. We observe substantially less heterogeneity in $_{it}^{price}$ and we also observe that the MTE and ATE measures tend to be more similar to one another.

A.3 Discrepancy Between Average and Marginal Treatment E ects

We perform a Monte Carlo study to analyze the extent to which the average treatment e ect deviates from the marginal treatment e ect under di erent demand speci cations. We generate data by simulating from a random coe cients logit model with a single random coe cient on price. Our simulations follow the procedure in Armstrong (2016), Judd and Skrainka (2011) and Conlon (2016), in which prices are endogenously determined via a Bertrand-Nash game given the other utility parameters (rather than directly drawn from a distribution).

We generate the data in the following manner: $u_{it} = {}_0 + x_j \, {}_1 \quad {}_i p_j + {}_j + "_{ij}$ and $mc_j = {}_0 + {}_1 x_j + {}_2 z_j + {}_j$ where $x_j ; z_j \quad N(0;1)$, with $_j = ! \, {}_{j1} + (1 \quad)! \, {}_{j2} \; j \; 1 = ! \, {}_{j1 + (1} + \quad)! \, {}_{j}^{2}$

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

Table 14: Alternative Speci cations for Nevo (2000).

	-1.000	-1.000	-1.000	-1.000	-4.000	-4.000	-4.000	-4.000

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).[59] Because some potential substitutes see q_k 0 and may have priors s_k near zero, we need to bound the prior probabilities away from zero in order to avoid drawing from degenerate distributions. Therefore we add 1.1 pseudo observations from a uniform prior $\frac{1}{K+1}$ to each substitute. This gives a Dirichlet parameter of $_k = \frac{s_k}{1}$

Experiment	Dirichlet	Dirichlet	Normal-Logit
Prior Mean $_j$ =	$\frac{s_k}{1 \quad s_j}$, (s_j = 0:75)	$\frac{1}{K+1}$	$\frac{1}{K+1}$

+1

Mfg	Product	Treated Machine Weeks	Avg # Cntl Per Trt	q_k Subst Sales	q_j Focal Sales	q_k = j q_j j Div	Treated Machine Weeks	Avg # Cntl Per Trt	q_k Subst Sales	q_j Focal Sales	q_k = j q_j j Div
Nestle	Nonchoc Nestle (Con)	6	80.3	14.1	-19.8	71.1	3	8.7	9.4	-10.5	89.5
Mars	M&M Peanut	186	120.3	482.4	-915.9	52.7	176	10.0	375.5	-954.3	39.4
Mars	Twix Caramel	143	120.3	339.6	-682.6	49.7	134	9.8	289.6	-702.4	41.2
Misc	Farleys (Con)	22	40.9	41.0	-121.2	33.8	18	4.6	14.9	-114.2	13.0
Hershey	Choc Hershey (Con)	51	51.9	62.1	-210.0	29.6	41	8.8	29.8	-179.6	16.6
Mars	M&M Milk Chocolate	104	116.1	114.7	-454.6	25.2	97	10.6	71.8	-457.4	15.7

Manuf	Product	Focal Sales	No Prior	Beta-Bin Diversion m = J[y]	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	Butter nger	-362.8	20.1	17.1	14.3	11.2	7.8
Mars	Choc Mars (Con)	-32.7	19.7	6.5	3.5	2.0	1.0
Pepsi	Rold Gold (Con)	-900.1	17.9	16.8	15.7	13.9	11.6
Hershey	Choc Hershey (Con)	-179.6	16.6	12.2	9.1	6.3	3.9
			Zoo Animal Crackers Removal				
Hershey	Payday	-0.4	84.7	0.6	0.3	0.1	0.1
Kellogg	Rice Krispies Treats	-37.8	62.2	23.2	12.7	7.2	3.9
Misc	Salty United (Con)	-18.9	55.1	12.6	6.3	3.4	1.8
Kraft	Oreo Thin Crisps	-37.8	39.4	14.7	8.0	4.5	2.4
Pepsi	Rold Gold (Con)	-440.8	25.9	22.9	19.8	16.2	12.1
Hershey	Choc Hershey (Con)	-132.6	25.3	17.1	12.0	7.9	4.7
Misc	Hostess Pastry	-62.2	23.7	11.8	7.2	4.4	2.5
Kraft	Chse Nips Crisps	-37.8	23.1	8.6	4.7	2.6	1.4
			Chocolate Chip Famous Amos Removal				
Nestle	Choc Nestle (Con)	-0.2	300.0	1.2	0.6	0.3	0.2
Hershey	Choc Hershey (Con)	-66.8	72.7	36.9	22.5	13.4	7.4
Kraft	Oreo Thin Crisps	-43.3	47.9	19.2	10.8	6.1	3.3
Pepsi	Sun Chip	-355.7	40.4	34.4	28.9	22.7	16.1
Hershey	Payday	6.8	38.9				
Misc	Salty United (Con)	-28.7	34.6	10.7	5.6	3.1	1.7
Pepsi	Chs PB Frito Cracker	-83.6	32.1	18.2	11.6	7.1	4.1
Kraft	Planters (Con)	-332.6	24.7	20.9	17.5	13.7	9.8
			M&M Peanut Removal				
Misc	Hostess Pastry	-38.6	32.5	12.3	6.9	4.0	2.3
Mars	Snickers	-1239.3	23.9	22.9	21.7	19.9	17.2
Kellogg	Brown Sug Pop-Tarts	-43.5	22.9	9.2	5.2	2.9	1.6
Misc	Cli (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 19: Sensitivity of Beta-Binomial Diversion to Number of Pseudo Observations

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

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 di erent number of pseudo-observations.

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

A.6 Stan Code for MCMC Estimator

This is code for the R library stan (Team 2015) which recovers the MCMC estimator of the