

Effects of Product Availability: Experimental Evidence ^{*}

Christopher T. Conlon[†]
Julie Holland Mortimer[‡]

July 18, 2011

Abstract

Product availability impacts many markets in which stochastic demand and fixed inventories may lead to stock-outs. In vertically-separated markets, optimal inventory decisions may differ substantially between upstream and downstream firms, making vertical arrangements necessary for coordination. We examine the impact of exogenously removing products from vending machines, and analyze subsequent changes in purchasing patterns and profits using both nonparametric analyses and structural demand estimation. We find substantial switching to alternate products, and evidence of misaligned stocking incentives between upstream and downstream firms. We discuss the trade-offs of both empirical approaches for analyzing product availability effects generally.

^{*}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 Dan Akerberg, Kate Ho, Greg Lewis, Richard Mortimer, and seminar participants at Harvard University and UCLA for helpful comments. Tom Gole, Adam Kapor 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.

[†]Department of Economics, Columbia University, 420 W. 118th St., New York City, NY 10027. email: cc3264@columbia.edu

[‡]Department of Economics, Harvard University, 1805 Cambridge St., Cambridge, MA, 02138, and NBER. email: mortimer@fas.harvard.edu

1 Introduction

Product availability impacts many markets, particularly those for which storage costs or capacity constraints matter. For example, inventory decisions may lead to stock-outs in retail markets, and capacity decisions may affect transportation, performance events, health care provision, and school choice. Firms in these markets may optimize product availability to influence consumer decisions about where and when to shop. In vertically-separated markets, optimal stocking choices for downstream firms may differ substantially from those of upstream manufacturers. In such settings, manufacturers may use vertical arrangements to try to align the stocking decisions of the downstream firms with their own interests.

Despite the key role that product availability plays in many markets, little empirical evidence exists on the importance of product availability for firms or consumers. We investigate the impact of product availability through a field experiment in which we exogenously remove top-selling products from a set of vending machines, and track subsequent consumer responses and profit impacts. Product availability in the context of vending has been the subject of many recent debates about obesity and the appropriate public policy response to the mix of products offered in vending machines, particularly in school settings.¹

We find that most consumers purchase another good when a top-selling product is removed. The profit impacts indicate that stocking incentives may be misaligned between downstream and upstream firms in the absence of vertical arrangements. For example, some product removals result in lower revenues for the upstream manufacturer but higher profits for the downstream firm as consumers substitute to products with higher downstream margins. This effect may provide a rationale for the use of tying or forcing contracts by upstream firms that is unrelated to competitive considerations at the upstream level.

The experimental nature of the data allows us to analyze the substitution patterns and profit impacts of each product removal using two alternative methodologies. First, we report nonparametric analyses of the data using techniques common to the treatment-effects literature, and applicable here due to the exogenous variation in product availability introduced by the experiment. Second, we analyze the data using structural demand estimation. The implementation of the two estimation methods allows us to generate insights into the relative benefits and drawbacks of applying these alternative methodological approaches to other data sources or in other settings.

We run our field experiment on a group of 60 vending machines located in five office buildings in downtown Chicago. Over the course of a three-year period, we implement six single-product removals (in which a single product is removed from all sites), and two double-

¹Forty states now tax junk food or soda products, and cities, school districts, and other local jurisdictions have proposed or implemented restrictions on the set of products that may be offered in vending machines. See Engber (2009) for a recent press article summarizing many policy responses in this area. More recent examples include rules requiring that the mix of beverages in city vending machines favor water in New York City, a ban on sales of sugary drinks in city buildings in San Francisco, and a similar proposed ban in Boston (Smith 2010). The medical literature has also weighed in on the issue of taxing sugary drinks (e.g., see Brownell and Frieden (2009) and Brownell, Farley, Willett, Popkin, Chaloupka, Thompson, and Ludwig (2009)).

product removals (in which two products are removed simultaneously). Each removal lasts for 2-3 weeks. For the nonparametric analyses, we choose control weeks from the three-year panel using nearest-neighbor matching methods and report the substitution patterns implied by a comparison of the treatment and control weeks. Using the same control weeks, we also report the profit impacts of each experiment for the downstream firm and the revenue impacts for all upstream firms.

The nonparametric approach allows us to document changes in purchasing patterns and profit impacts directly, with no need to make any parametric assumptions about the functional form of demand. The changes in purchase frequencies across remaining products are quite sensible; most consumers switch to a similar product when their first choice is removed, and relatively few consumers walk away. In some cases, the downstream firm appears to profit from a product removal, because consumers switch to products with higher downstream profit margins. However, accounting for the vertical contracts between the vending operator and upstream manufacturers reverses this effect, indicating that one rationale for the vertical arrangements in this industry may be to align the downstream firm’s incentive to carry and service the products of upstream manufacturers.

The nonparametric approach requires no functional form assumptions, but this flexibility can produce noisy measures of the true effect of each product removal. For example, total vends are occasionally higher during the treatment period (when a product is removed) than during the control period, and vends of substitute goods are sometimes lower during the treatment period than in the control. We would not want to assign a causal interpretation to these types of outcomes, and thus, they represent a limitation to the otherwise very informative results of this approach. More broadly, these sorts of problems may be endemic to any large-scale field experiment in industrial organization (IO). Unlike studies of individual behavior in the lab or the field, randomization may not be able to control for all possible market-level variables in most IO contexts. For example, we can’t prevent Mars, Inc. from advertising Snickers in different ways over time or in different geographic markets.

The second way in which we analyze the substitution patterns and profit impacts of changes in product availability is through the use of structural demand estimation. Using the full dataset, we estimate nested logit and random-coefficients logit models of demand. We predict vends during the treatment and control periods using the estimated model parameters, and compare these outcomes. The models perform well in many respects, and capture much of the variation that is observed in the nonparametric results. However, they tend to predict fewer sales to the products identified as ‘top substitutes’ in the nonparametric analysis, and predict more sales to other products. We speculate on two reasons for this result; namely, properties of the logit error term, and the endogeneity of changes in the retailer’s product mix. We also provide estimates of downstream profit and upstream revenue, and again find evidence of mismatched incentives in the vertical supply chain that are addressed by the vertical contracts used between this vending operator and the upstream manufacturers.

Next, we extend the structural analyses by simulating the effect of the product removal directly. This differs from the first exercise, in which we compare estimates of the model

during the treatment and control periods, because that exercise allows for other changes to occur in addition to the stock-out of the focal product. For example, if a different brand of pretzels was carried during a control week, the first exercise adjusts for that. In this sense, neither the nonparametric analyses nor the predicted structural estimates of vends during the treatment and control periods provide a clean measure of the impact of the product removal *per se*. Simulating the product removal, on the other hand, holds all other factors fixed, and isolates the “pure” treatment effect when the model is correctly specified. The results of simulating the product removal are qualitatively similar to the ‘prediction’ exercise. The profit(revenue) impacts for downstream(upstream) firms are quite similar to the impacts we estimate using the nonparametric approach, with the same instances of mismatched incentives in the vertical chain.

In most cases in which one estimates demand, the type of exogenous variation in product availability that we create through our experiments does not exist in the data. Thus, structural demand models are often identified solely from naturally-occurring variation in choice sets (which may also include variation in product characteristics, such as price). One way to examine how successful these sources of variation are for identifying model parameters is to remove the variation in choice sets that arises from our experimental interventions. Thus, we conduct a series of hold-out analyses, in which we estimate the nested logit and random-coefficients logit models using subsets of the data that drop treatment periods.² We find that the estimated model parameters are generally robust to the exclusion of data from the experimental interventions, with the exception of the parameters that govern correlation in consumer tastes for candy and sugar. These parameters are sensitive to the exclusion of data from the candy bar stock outs. The finding is intuitive because there are fewer “naturally occurring” changes over time in the product mix of the candy category, and the candy bars that we exogenously remove have much higher market shares than the salty snack and cookie products that we exogenously remove.

There are several advantages to studying product availability in the context of vending machines. One advantage is the ability to successfully implement the field experiments logistically (ie., to exogenously change the set of products that are available to consumers). Even a “simple” retail setting would introduce complications to this change that are absent with vending machines.³ Second, the scale of revenues that are potentially at stake in a vending machine is relatively small, so the experiments are not prohibitively expensive to run. Third, we observe the wholesale costs of the vending operator, which makes calculation of the upstream and downstream revenue/profit impacts possible.

Some features of vending machines are both advantages and disadvantages for the purpose of our study. Competition between retail outlets is not a feature of our setting. While this simplifies many aspects of the study, it is also a limitation, in the sense that we cannot

²One could describe the baseline estimates from our full dataset as an “in-sample” prediction exercise for the structural models. The hold-out analyses provide an “out-of-sample” prediction, in which the model is asked to predict the results of experiments that have not already occurred in the data.

³For example, one would need to remove the focal product not only from shelves, but also from back-room storage areas, or alternatively prevent clerks from responding to special consumer requests to retrieve an item from storage when it is not on the shelf.

study how competition among downstream firms responds to changes in product availability. Similarly, price variation is quite limited, especially within a site for a category (e.g., candy) over a period of time.

Finally, we note that the field experiment is not a randomized trial. Rather, each removal exogenously varies the set of products available to all consumers for a period of time, and control weeks are selected from an observational dataset during which no intervention occurred.⁴ We did not have the ability to randomize the set of products offered to any given consumer at the time of her purchase. Such an experiment would represent a true randomized trial (absent other market-level effects), but is not feasible technologically in brick-and-mortar retail contexts.

Relationship to Literature

This paper connects several different literatures. The first is a growing literature in economics, marketing, and operations research that focuses on firms’ stocking decisions and the importance of product availability for vertical arrangements. The “newsvendor” problem dates back to Edgeworth (1888), Spengler (1950), and Arrow, Harris, and Marschak (1951), and describes the potential for mis-aligned inventory incentives between upstream and downstream firms. More recent theoretical work formalizes and extends the solution to this problem (e.g., Kraiselburd, Narayanan, and Raman (2004), Schweitzer and Cachon (2000), and many others). In more recent empirical work, Anupindi, Dada, and Gupta (1998) study product availability, also in the context of vending machines. Several examples in this literature focus on scanner data and availability at supermarkets and convenience stores, such as Bruno and Vilcassim (2008), and Musalem, Olivares, Bradlow, Terwiesch, and Corsten (2010), and Matsa (2010). Aguirregabiria (1999) uses scanner data to examine the strategic implications of dynamic inventory decisions in the context of vertically-separated markets.

The second is a wider literature on field experiments in economics. A rather extensive review of this literature is presented in Levitt and List (2009). Some recent and notable examples include Karlan and List (2007) who study the impact of price on charitable giving, and Bertrand, Karlan, Mullainathan, Shafir, and Zinman (2010) who examine the impact of advertising using a direct-mail experiment involving consumer lending. Much of the field experiments literature focuses on direct-mail, charitable giving, or auction settings (such as Engelbrecht-Wiggans, List, and Reiley (2005) or Ostrovsky and Schwarz (2009)). In a retail setting Cai, Fang, and Yuyu (2009) examine observational learning by randomly marking menu items as “favorites” and analyzing the impact of the designation on customer demand. There is a small related literature that uses experiments to study the effects of stockouts. Fitzsimons (2000) studies psychological effects of stockouts on consumers in the laboratory,

⁴From the point of view of experimental economics, our intervention is more closely related to laboratory experiments that attempt to understand markets as a whole (e.g. Smith (1962) and more recently, Eriv and Roth (1998), among others), than to experiments that examine individual responses (Tversky and Kahneman (1991) summarize many examples). Despite the fact that such interventions fall short of a fully randomized trial, they are in fact exactly the type of experimentation that one might expect retailers to do when trying to learn about demand or set prices in an optimal way. Online retailers may differ from brick-and-mortar stores in this respect, but the nature of competition and consumer search may also differ in online settings.

and Anderson, Fitzsimons, and Simester (2006) examine psychological framing effects of how stockouts are presented to consumers in the context of a mail-order company.

The paper also contributes to a recent discussion about the role of different methods in empirical work going back to Leamer (1983), and discussed recently by Heckman (2010), Angrist and Pischke (2010), Leamer (2010), Keane (2010), Sims (2010), Nevo and Whinston (2010), Stock (2010), and Einav and Levin (2010). A central issue in this debate is what role experimental or quasi-experimental methods should play in empirical economic analyses in contrast to structural methods. Several of these recent papers essentially argue that both types of approaches have advantages and drawbacks. Our setting provides the opportunity to examine empirically the trade-offs to which these papers refer. For example, while our experimental estimates are quite informative in many respects, there are cases in which we would not want to infer causality (e.g., when overall sales increase during a stock-out event). The structural demand models use economic theory to rule out such an effect, but cannot fully capture the degree of substitution that occurs from a focal product to other goods. This is especially true when a product’s most important characteristics are less easy to measure (e.g., packaging differences, or possibly unobserved advertising campaigns).

The paper proceeds as follows. We describe the vertical arrangements used in the vending industry, and the design of our field experiments and data in section 2. Section 3 describes the non-parametric results from the field experiments including the implications of the experimental results for firm profitability. In section 4, we describe two structural demand models commonly used to estimate substitution patterns (i.e., nested-logit and random-coefficients logit models), and a method for estimating the rate of consumer arrivals at each machine. In section 5 we compare the predicted substitution patterns of those models during the treatment and control weeks analyzed in section 3; section 6 uses the demand models to simulate the removal of focal products (holding all other conditions of the market fixed). Section 7 provides the results of hold-out analyses, in which we estimate the demand model on subsets of the data, and section 8 concludes.

2 The Vending Industry and Experimental Data

2.1 Vertical Arrangements in the Vending Industry

Vertical arrangements are widely used in the vending industry and apply to several of the upstream relationships of the firm with whom we worked. The most commonly used vertical arrangement in the industry is referred to as a “rebate program.” Under a rebate program, a manufacturer refunds a portion of a vending operator’s wholesale cost at the end of a fiscal year if the vending operator meets an annual sales goal, typically expressed as a percentage of last year’s sales. The sales goal for an operator is typically set for the combined sales of a manufacturer’s products, rather than for individual products. Some manufacturers also require a minimum number of product “facings” in an operator’s machines. The amount of the rebate and the precise threshold of the sales goal or facing requirement is specific to an individual vending operator, and these terms are closely guarded by participants in the industry.

While the rebate programs share features with vertical tying (particularly when facing requirements apply), the primary function of the rebate programs is to lower the wholesale cost through an operator-specific mechanism.⁵ The benefit of rebate programs for manufacturers is the ability to more closely align the downstream operator’s incentive to carry and re-stock the manufacturer’s products with the manufacturer’s own incentives. Specifically, at any wholesale cost greater than the cost of production, the downstream firm chooses to stock fewer units of inventory than the upstream manufacturer would choose. This inventory stocking problem is well understood, and is referred to as the “newsvendor” problem in the case when prices are fixed and demand is stochastic. The intuition is formalized as follows.

Consider a single product with stochastic demand denoted by D , with distribution function F . The downstream vendor purchases q units of inventory at cost c and sells at a fixed price $p > c$. Demand lasts for one period, and inventory is purchased at the beginning of the period. For the sake of illustration, assume no salvage value for the downstream firm and a production cost for the upstream firm of zero.⁶ Realized profit is:

$$\pi(q, D) = p \min(q, D) - cq \quad (1)$$

and expected profit is:

$$E[\pi(q, D)] = (1 - F(q))\pi(q, q) + \int_0^q f(x)\pi(q, x)dx. \quad (2)$$

The first term captures profits when the firm is understocked (i.e., no incremental profit is earned from consumers who arrive after the product is sold out), and the second term captures profits when the firm is overstocked. The solution to the newsvendor problem maximizes expected profit at quantity:

$$q^* = F^{-1}\left(\frac{p - c}{p}\right) \quad (3)$$

Thus, q^* is chosen so that the expected marginal return equals the marginal cost to the downstream firm of an additional unit of inventory.⁷ Note that the newsvendor problem is analogous to a “fixed price/stochastic demand” version of the double-marginalization problem, in that the downstream vendor only accounts for his own mark-up ($p - c$) rather than the full difference between p and production cost (assumed here to be zero) when stocking inventory.

Rebates lower the wholesale cost for downstream operators, leaving them with a higher expected return from stocking an additional unit of inventory. By structuring this as a

⁵Robinson-Patman prevents manufacturers from directly price discriminating across competing downstream firms when selling “inputs.”

⁶This is without loss of generality for a production cost and/or salvage value for the downstream firm less than c .

⁷Alternatively, this ratio (referred to as the “critical fractile”) may be viewed as relating the cost of understocking (a lost sale is worth $(p - c)$) to the total cost of being either understocked or overstocked (i.e., the cost of understocking $(p - c)$ plus the cost of overstocking (c)).

rebate, rather than directly reducing wholesale price, manufacturers are able to tailor the amount of the cost reduction to each individual operator, and to match it to targets that are retailer specific (e.g., 90 percent of his previous year’s sales).

2.2 Experimental Design

We ran eight experimental treatments with the help of Mark Vend Company, which is a medium-sized independent vending operator in the Chicago area. We identified 60 snack machines located in office buildings, for which demand was historically quite stable.⁸ Most of the customers at these sites are ‘white-collar’ employees of law firms and insurance companies. Our goal in selecting the machines was to choose machines that could be analyzed together, in order to be able to run each experiment over a shorter period of time across more machines.⁹ We selected snack machines because beverage machines have extremely large capacities and a small number of products. This made the logistics of stocking out beverages more difficult (removing and storing 100 large heavy bottles vs. 20 candy bars), and also made the outcomes less interesting, because the demand system only includes around six products. Finally, we selected machines on routes that were staffed by experienced drivers, so that the implementation of the experiments would be successful. The 60 machines used for each experiment were distributed across five of Mark Vend’s clients, which had between 3 and 21 machines each. The largest client had two sets of floors serviced on different days, and we divided this client into two sites. Generally, each site is spread across multiple floors in a single high-rise office building, with machines located on each floor.

Implementation of each product removal was fairly straightforward; we removed either one or two top-selling products from all machines for a period of roughly 2.5 to 3 weeks. Six of the experiments stocked-out a single top-selling product: Snickers, Peanut M&Ms, Zoo Animal Crackers, Famous Amos Chocolate Chip cookies, Doritos, or Cheetos. Two of the experiments removed two products simultaneously: Snickers plus Peanut M&Ms, or Doritos plus Cheetos. Whenever a product was experimentally stocked-out, poster-card announcements were placed at the front of the empty product column. The announcements read “This product is temporarily unavailable. We apologize for any inconvenience.” The purpose of the card was two-fold: first, we wanted to avoid dynamic effects on sales as much as possible, and second, the firm wanted to minimize the number of phone calls received in response to the stock-out events.

The dates of the interventions range from June 2007 to September 2008, with all removals run during the months of May - October. We collected data for all machines for just over three years, from January of 2006 until February of 2009. During each 2-3 week experimental period, most machines receive service visits about three times. However, the length of service visits varies across machines, with some machines visited more frequently than others.

⁸More precisely, demand at these sites is “relatively” stable compared to the population of sites serviced by the vending operator.

⁹Many high-volume machines are located in public areas (e.g., museums or hospitals), and have demand that varies enormously from one day to the next, so we did not use machines of this nature. In contrast, the work-force populations at our experimental sites are relatively homogenous.

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. With the exception of an individual site on each of two experimental runs, implementation was successful.¹⁰

We faced a few limitations when designing the experiment. For example, some removals were scheduled “back-to-back.” In these cases, we selected products that seemed ex-ante less likely to be close substitutes for adjacent runs. For example, the Doritos stock-out was followed by the Peanut M&Ms stockout. Due to more difficult logistics associated with experimental price changes, we were not able to implement any pricing experiments.¹¹ Finally, throughout our analyses, we focus on static effects. We do not see much evidence of dynamic effects in the data, but this is not something for which we are able to test directly. We note that demand for a focal product tends to remain fairly stable (and demand for other products returns to previous levels) after it is replaced.

2.3 Data Description

Data on the number and price of all products vended are recorded internally at each vending machine used in our experiments. The data track vends and revenues since the last service visit (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 prices and variable costs for each product at each service visit during our 38-month panel. There is relatively little price variation within a site, and almost no price variation within a category (e.g., candy) at a site. Very few “natural” stock-outs occur at our set of machines.¹² Over all sites and months, we observe 162 unique manufacturer products. We organize these products into 417 site-product pairs (approximately 70 unique manufacturer products per site) by consolidating low-selling products over time within each site.¹³ This set of 417 site-product pairs is our base dataset for all analyses and estimation.

In addition to the data from Mark Vend, we also collect data on the characteristics of

¹⁰In the two unsuccessful runs, the driver at one site forgot to remove the focal product, so no intervention took place.

¹¹The firm does change some prices at some sites late in the dataset; we do not analyze this variation.

¹²Mark Vend commits to a low level of stock-out events in its service contracts.

¹³For example, we combine Milky Way Midnight with Milky Way. In a small number of cases, the consolidated products vary slightly in their wholesale costs or combine products of different manufacturers. For these products, we use the modal wholesale cost, and we apportion revenues paid by the vending operator to manufacturers on the basis of the vends of each individual product.

each product online and through industry trade sources.¹⁴ For each product, we note its manufacturer, as well as the following set of product characteristics: package size, number of servings, and nutritional information.¹⁵ Summary statistics at the manufacturer level are reported in the Appendix. One variable that the data do not measure is the number of people who walk away from a machine. We considered the possibility of adding video cameras or pressure mats to the machines, but neither of these options would have provided clean information on market size.¹⁶ We discuss the issue of market size in detail when we describe the structural models of demand.

3 Nonparametric Analyses of the Experimental Outcomes

3.1 The Matching Estimator

In order to calculate changes in purchasing patterns, sales during treatment weeks are compared with sales during control weeks. We measure substitution from product k to product j as:

$$\Delta q_j = E[q_j | A_{J \setminus k}] - E[q_j | A_J]$$

where q_j denotes weekly sales, J is the full set of products, and A_J denotes availability of all products in J .

In principle, this calculation is straightforward. In practice, however, there are three challenges in implementing the experiments and interpreting the data generated by them. First, service visits vary in length across machines and over time. Second, overall sales levels vary over time, due to exogenous changes in the rate of consumer arrivals. For example, a law firm may have a large case going to trial in a given month, and vend levels will increase at the firm during that period. Third, the product mix presented in a machine is not necessarily fixed across machines, or within a machine over long periods of time (e.g., as manufacturers change their product lines, or Mark Vend changes stocking decisions). Variation in the product mix across sites and machines increases the number of outcomes that the experiment attempts to measure (consider that we start with 162 unique products, roughly 70 of which are carried at any particular site).¹⁷ Changes in the product mix that occur over time for a given machine affect the comparability of the observational control weeks to the weeks in which treatment occurs.

¹⁴For 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.

¹⁵Nutritional information includes weight, calories, fat calories, sodium, fiber, sugars, protein, carbohydrates, and cholesterol.

¹⁶Pressure mats were not workable because potential customers can see the product facings without standing close enough to be registered on the mat. Video cameras would have introduced issues of human subjects approval into the experiments, and also suffer from the problem of consumers being able to see the product facings in some machines without standing close enough to show up on the video.

¹⁷Recall that a single machine may be able to stock roughly 35 different products at one time, depending on configuration.

We take three steps to address these complicating factors. First, we consolidate data collected at service visits to weekly observations. This allows us to make direct comparisons across machines that are visited at different frequencies.¹⁸ Second, we create classes of products for reporting the results of the experiments. This reduces the number of outcomes that each experiment attempts to measure. Third, we select a set of matched control weeks using nearest-neighbor matching techniques. This adjusts for the fact that the treatment weeks may belong to periods of low or high demand, due to exogenous variation in market size over time. We describe each of these steps in turn.

The first step for analyzing the experimental outcomes is the assignment of sales data to weekly units. This is done by apportioning “total vends since the previous visit” evenly across the elapsed days, and gathering groups of seven days into weeks.¹⁹ Moving to weekly visits has a number of advantages. It smooths out variation in sales levels that occurs over very short (daily) intervals, and makes data from different service visits comparable.²⁰ Experiments are implemented at service visits, and all data from the experiments are also apportioned to weeks. This introduces the possibility that an experiment may be “contaminated” at the weekly level because implementation of the stock-out occurred in the middle of a week. In order to minimize this contamination, we choose start-dates for weeks at each site based on the timing of the experimental stock-out service visit.²¹ This eliminates contamination, except for a very small number of cases in which different machines within a site are visited on different days. In such cases, we eliminate any treatment weeks for which more than four vends of the focal product were recorded at the site.

The second step we take for the analysis is to create product “classes” by combining the roughly 70 products at each site into groupings for which changes in purchase frequencies can be reliably measured. This is necessary because smaller products are not bought in sufficient quantity to identify changes for each one individually.²² We allow the product classes to vary by experiment: all experiments include the six focal products, four additional major products (Twix Caramel, Salted Peanuts, Raisinets, and Skittles), and seven “assorted” classes to capture the smallest products. We choose several additional individual products to track for each product removal based on the purchase patterns observed in the data.²³

¹⁸When we estimate models of demand in later sections, we preserve the visit-level detail for estimation, but report predictions and simulations of the model at the weekly level.

¹⁹On average across the entire dataset, service visits are approximately 2 days long. However, the time between service visits ranges from one to 21 days. The length of time between visits varies over time for a given machine, and also across machines (and sites).

²⁰Daily vends would also provide comparable results, but with a significant amount of noise in overall levels of vends.

²¹For example, if Snickers bars were removed on a Wednesday at site 93, we define weeks as periods of Wednesday - Tuesday at that site. We set the start-date of weeks separately for each site and each experiment, and require that each week start on a workday (Monday - Friday).

²²Fig Newton and Swiss Creme cookies are examples of these types of products. The stocking of these products is sometimes governed by the client (e.g., the CEO may request that Fig Newtons be available).

²³The assorted classes are: Chocolate, Candy, Energy, Cookie, Potato Chips, Pretzels, and Salty Snacks. As an example of how the additional individual products vary across experiments, consider the Snickers and Cheetos removals. When stocking out Snickers, the additional individual products are Ruger Vanilla Wafer and Grandma’s Chocolate Chip Cookie. For the Cheetos stock-out, the additional individual products are

The third step is to select matched control weeks from the three-year panel of observational data. Before selecting control weeks, we sum vends across the machines at a given site.²⁴ Levels of demand at a site that change over time affect our ability to compare sales during treatment and control periods, and the selection of matched control weeks focuses on choosing weeks in which the level of demand is similar to that during the relevant treatment period. In our particular setting, many of the experiments were run during the summer of 2007, which was a high-point in demand at these sites, most likely due to macroeconomic conditions. In order to select weeks of similar demand levels, we identify a set of product classes that we believe are ex-ante unlikely to be substitutes to the focal product, and we use nearest-neighbor matching methods, matching on the site-level sales of these “non-substitute” product classes.

The use of matching on non-substitute product classes may be motivated in the following way. Substitution from product k to product j in market t is described as the change in the probability of purchasing j when k is not available. Excluding cases in which products are complementary in consumption, this implies that

$$p_{jt}(A_{J \setminus k}) \geq p_{jt}(A_J).$$

In the data, we observe sales rather than choice probabilities, given by:

$$q_{jt} = M_t p_{jt}(A_t),$$

where A_t denotes the set of available products in market t , and M_t denotes market size.²⁵ The challenge for identifying substitution is that M_t is unobserved. The matching procedure attempts to control for changes in M_t by matching on sales levels of non-substitute products. For non-substitute product l ,

$$p_{lt}(A_{J \setminus k}) \approx p_{lt}(A_J).$$

Thus,

$$\frac{q_{lt'}(A_{t \setminus k})}{q_{lt}(A_t)} \approx \frac{M_{t'}}{M_t},$$

so that by matching on sales levels of the non-substitute goods, we try to obtain a ratio of $M_{t'}/M_t$ that is close to one.²⁶

Sun Chips, Frito, Farley’s Mixed Fruit, Cheez-It Original, and Hot Stuff Jays.

²⁴There is nothing to be gained in the nonparametric analysis from variation across different machines at the same office building, given that our experiment removes each focal product from all machines at the site at the same time.

²⁵A market in our context generally has both geographic and temporal characteristics (ie., the week of March 4th at site 93).

²⁶We investigated a number of alternative approaches. For example, we analyzed a “difference-in-difference” type estimator in which, instead of matching the treatment week to a set of control weeks, one matches the week before an experiment to potential control weeks on the basis of focal-product sales and calculates changes in vends over time. This avoids the problem of having to choose products on which

Matching is done within each site. For each treatment week we select the four closest control weeks based on sales of the non-substitute product classes.²⁷ The set of products that are used for matching are shown in table 16 of the Appendix. We grouped the salty snack experiments together, and the candy and Chocolate Chip Famous Amos experiments together for defining sets of products for matching.²⁸ We use different subsets of products on which to match at different sites, due to changes in availability or the product mix of the assorted classes at particular sites.

These three steps enable us to examine the results of each experiment. Note that for each experiment, we have one outcome for each non-focal product class at a site. Matching estimators are usually discussed in the context of a single outcome of interest, such as in Lalonde (1986), or Dehejia and Wahba (1999).²⁹ In our context, the “average treatment effect” is a vector of outcomes because we have multiple outcomes of interest. Our approach, therefore, is to use the matching methods developed in the treatment effects literature to generate matched observations.³⁰ We then use this matched sample to report mean outcomes for all products for the treatment and control weeks, along with the percentile of the distribution of all control outcomes to which the mean results relate. Thus, we measure substitution from product k to product j as:

$$\Delta q_j = E_i[q_{ij}|A_{J \setminus k}, T_i = 1] - E_i[q_{ij}|A_J, T_i = 0, N_i \leq 4]$$

where expectations are taken over weeks, indexed by i , and J denotes the full set of product classes. The variable T_i is an indicator variable denoting whether week i belongs to the treatment period in which product k was exogenous removed, and N_i is the distance rank

to match based on ex-ante notions of substitutability. In practice, however, week-to-week (or visit-to-visit) variation is quite noisy, and this resulted in difference-in-difference estimators that primarily captured random week-to-week fluctuations. We also ran all analyses using a set of “admissible” control weeks that were identified on the basis of a site’s product mix during the experimental period. In this method, we admitted control weeks for which each product carried at a site during the experimental weeks was available in at least 80 percent of the machines at each site. This method yielded qualitatively similar results to the matching estimates that we report, but resulted in more experiments for which total vends increased when a product was removed. We report the nearest-neighbor matching estimator (using non-substitute products as the matching variables) as our baseline estimates because the statistical properties of these estimators are well understood. (See Abadie and Imbens (2006).) In contrast, admitting control weeks on the basis of a non-linear function of a vector of availability dummies for a set of products is (to us) less well understood. Finally, we also ran all analyses using the full control set of approximately 120 weeks. The matched estimators perform significantly better, particularly with respect to levels of total vends.

²⁷All estimates were also run using ten matched control weeks for each treatment week; results were qualitatively similar to the baseline estimates reported here for four matches.

²⁸The Chocolate Chip Famous Amos cookie experiment was grouped with the candy experiments because of the presence of chocolate. The set of products for matching in the Zoo Animal Cracker experiment was allowed to differ from the candy and Famous Amos experiments because the vending operator identified that product *ex-ante* as potentially having a different set of substitute products.

²⁹Abadie and Imbens (2006) work out the large sample properties of matching estimators for average treatment effects in this context, and Imbens (2004) provides a review of this literature.

³⁰We use the `nnmatch` command in Stata, described in Abadie, Drukker, Herr, and Imbens (2004), and choose the Mahalanobis metric for measuring the distance between the treatment and control vectors of covariates.

of each potential control week produced by the matching procedure.³¹ We also estimate the effect of a product removal on the inside market shares of all remaining products. Thus, we measure:

$$\Delta s_j = E_i \left[\frac{(q_j | A_{J \setminus k}, T_i = 1)}{(\sum_j q_j | A_{J \setminus k}, T_i = 1)} \right] - E_i \left[\frac{(q_j | A_J, T_i = 0, N_i \leq 4)}{(\sum_j q_j | A_J, T_i = 0, N_i \leq 4)} \right]$$

where s_j denotes the inside market share of product j .

We estimate outcomes for each site and each experiment, which generates a set of 48 tables of outcomes (six sites times eight experiments). In order to capture the overall effect of an experiment, we sum over the average weekly rates at each site during the treatment and control periods, and compute the difference, as well as the percentile of the distribution of vends at all sites to which the rate corresponds.³² This adds eight more tables to the set of results.

3.1.1 Inference

Whenever we report the effect of a product removal, we report the quantile of the distribution of overall weekly sales for each outcome (treatment and matched control weeks). The quantile corresponding to the matched control weeks gives a sense of how the matched sample compares to the full distribution across all weeks. Similarly the quantile corresponding to the treatment outcome allows one to compare the outcome to the full distribution of weeks. However, neither statistic gives a direct way to infer statistical significance for the effect of the removal, according to the matching estimator. For this, we use a falsification study, similar to a bootstrap-type procedure. For each product removal, we choose three control weeks at random from each site, and assign these to be “treatment” weeks. For these false treatment weeks, we perform the same matching procedure, and construct the same outcomes of interest (vends, difference in vends, revenues, etc.), as in the baseline analysis. We repeat this for 1000 trials, and report the values of this falsification procedure at the 5th and 95th percentiles as a 90% Confidence Interval for the matching results.³³

³¹Thus, we denote the closest match as having rank $N_i = 1$, so that $N_i \leq 4$ selects the four nearest neighbors using the method in Abadie, Drukker, Herr, and Imbens (2004).

³²Note, this essentially weights the contribution of each site to the total impact of the experiment by its overall sales level. Different sites may run a particular experiment for shorter or longer periods relative to other sites. For example, one site may stock out a product for two weeks, while another stocks out the product for three weeks. An alternative weighting across sites would be to sum all vends together and then compute weekly rates. This would weight the average based on both sales levels at each site and the number of weeks for which a site ran a particular experiment. We prefer weighting by site because the length of time for a given experiment at a given site varies for unhelpful reasons. For example, slower sites may be visited less often and have longer runs but lower weekly sales.

³³Note that this procedure will sometimes result in a treatment mean that is outside the 90% confidence interval when the difference is not significant, and vice versa, depending on the variance in the matched control observations. When only one confidence interval is reported (e.g. tables 2 and 3), we report the significance level associated with the difference.

3.2 Changes in Purchasing Patterns

Table 1 reports results from one of the 56 tables—namely, the overall changes in purchasing patterns (summed over sites) of the Snickers removal.³⁴ The top panel of the table reports vends, and the bottom panel reports inside market shares. The first column in each table reports average weekly vends for the matched control weeks. The second column reports the percentile at which the mean of the matched control weeks falls relative to the full distribution of sales for all control weeks.³⁵ The third column reports average total vends for the treatment weeks, and the fourth column reports the percentile of the distribution of sales for all control weeks with which the treatment outcome is associated. If a product’s average total weekly sales during the treatment weeks exceeds total weekly sales for all control weeks, we report the 100th percentile. The fifth column reports the difference in the two means, and the last column reports the percentage increase in sales for the substitute good. For example, during the Snickers experiment, table 1 shows that Peanut M&Ms sold 118.4 more units in an average treatment week; its mean total weekly sales during the control period were 359.9, and the percentage increase was 32.9. Sales of Peanut M&Ms in the matched control sample exceeded sales in 73.6% of all control weeks, and the average treatment outcome of 478.3 sales exceeded sales in 99% of all control weeks. Both the treatment mean and the mean difference are outside the 90% Confidence Interval of the overall distribution for Peanut M&M sales. The magnitudes of the percentile changes among the ‘top substitutes’ are quite striking across all experiments, and it is common to see very large changes in sales percentiles between treatment and control periods for the top products.

Table 1 also shows that total vends during the matched control (treatment) period correspond to the 74th (73rd) percentile of the overall distribution of total vends. Overall, total vends are only 0.1% lower during the treatment weeks when Snickers is removed. This likely reflects at least two factors: first, most consumers purchase another product when Snickers is not available (as opposed to walking away), and second, demand was relatively high when the Snickers experiment was run. The rows in both panels are sorted by the mean difference in vends, so products toward the top of the list are those whose sales increased the most when Snickers was removed. Sales of the top five products (Peanut M&Ms, Twix Caramel, Assorted Chocolate, Assorted Energy, and Zoo Animal Cracker) increased by a total of 370 vends during the treatment period, which exceeds the average level of Snickers vends during the matched control weeks of 323. The products with the largest percentage change are found by examining the last column.³⁶ Examining the lower panel of table 1 allows one to normalize by overall sales levels by comparing changes in inside market shares.

Tables 2 and 3 summarize the results for all eight product removals. Table 2 reports the top five substitutes for each focal product(s) based on level changes in sales summed across

³⁴All 56 tables are available online in our online appendix. See the link for this paper at <http://mortimer.fas.harvard.edu>.

³⁵With no matching, most of these percentiles will be near 50%. The matched control weeks admits greater variance across products in these percentiles.

³⁶In the Snickers experiment, the top five products in terms of percentage changes are: Twix Caramel, Peanut M&Ms, Ruger Vanilla Wafer, Assorted Chocolate, and Zoo Animal Cracker.

sites; table 3 reports the top five substitutes for each focal product(s) based on percentage changes in sales. The last three rows in each panel report changes in the sales of: focal product(s), top 5 substitutes, and total vends between the treatment and matched control weeks. For all but one removal (Peanut M&Ms), we observe a reduction in total vends. Substitution to the top five products exceeds the average number of vends of the focal product during the matched control weeks in four removals in table 2.³⁷ The fact that vends to the top five substitutes exceed the vends of the focal product in this analysis implies that matching cannot fully control for changes in overall levels of demand across treatment and control weeks.

3.3 Profit Impacts

We observe prices and wholesale costs of Mark Vend, as well as its participation in rebate programs, so we can compute the total variable profit for each week with and without rebate payments. For upstream manufacturers, we observe revenue, but not production costs. We observe which products are co-owned by individual manufacturers, and we can calculate the revenue impacts across manufacturers that result from each product removal.

Table 4 reports the weekly profit impact of the Snickers removal for the downstream firm at the level of individual product classes, without manufacturer rebate payments, which are made annually based on Mark Vend’s total manufacturer-level sales.³⁸ Manufacturers are listed for each product class, with assorted product classes noting multiple manufacturers. The first column reports Mark Vend’s margin; his margin for Snickers (ignoring his rebate payments) is 21 cents, and he loses \$68.40 of variable profit per week on this product when it is removed. Sales of Assorted Salty Snacks are also down during the treatment week, and he loses \$51.60 of variable profit per week from this product class. He gains from other products: for example, increased sales of Peanut M&Ms contribute an additional \$25.20 in weekly profit. Overall, total vends are down, but the net profit impact for Mark Vend is positive, with an increase in profit of \$2.65 per week. This is generated in part by consumers’ willingness to purchase other products when Snickers is removed (rather than leaving empty-handed), and in part by the relatively low margin that Mark Vend receives on vends of Snickers (e.g., \$0.21 vs. \$0.48 for Assorted Energy products).

Table 5 summarizes the total weekly profit impact of each experiment for Mark Vend, with and without rebate payments. Two experiments (Snickers and Peanut M&Ms) result in a profit increase for Mark Vend when rebate payments are not included.³⁹ All three Mars & Co. product removals result in higher average margins for Mark Vend. The most striking example of this effect is seen for the double removal of Snickers and Peanut M&Ms, in which

³⁷Table 3 reports changes in the inside market share of the top five substitutes based on percentage changes. The combined change in the inside market shares of these five products exceeds the market share of the focal product in three experiments.

³⁸Products are listed in the same order as in table 1 (based on their increase in sales when Snickers is removed). The full set of 56 profit impact tables are provided in the online appendix.

³⁹However, the Peanut M&Ms result differs from the Snickers result because overall vends are higher during the treatment weeks than the control weeks in this case.

average margins increase by 1.2 cents per vend. The effect of this increase in margin is that Mark Vend loses only \$15.00 per week from the removal of these two products, despite a reduction in overall weekly sales levels of 218 units. The last two columns of table 5 report the change in Mark Vend’s average margin and profit impact for each removal, taking rebate payments into account. Mark Vend receives relatively generous rebate payments from Mars & Co., with the result that the profit impact of removing Snickers is negative, at -\$2.50, with his average downstream margin falling by 0.05 cents per unit, once rebate payments are considered.

Table 6 reports the revenue impact of each removal for all manufacturers, with and without rebate payments.⁴⁰ Revenue impacts for the manufacturer of the focal product of any given experiment are shown in bold typeface (e.g., Mars Inc. manufactures Snickers). Minor manufacturers include all 17 manufacturers listed in table 15. In all but one case, the manufacturer of the focal product had lower revenues during the treatment period in which its product was removed.⁴¹ The largest effect was for Mars Inc. during the joint stock-out of Snickers and Peanut M&Ms, for which its revenues declined by \$220.52 per week.⁴²

4 Structural Models of Demand

In this section we describe two models of demand commonly used for analyzing markets for differentiated products. Both models assume knowledge of market size, M . We first describe our method for determining market size, and then specify the structural demand models.

4.1 Calculating Market Size

Unlike many settings, in which market size is assumed to be constant over time, the number of people considering a purchase from a vending machine may change significantly from one week to the next. For example, the employees at one site may have an important deadline one week, which increases demand for vending products temporarily. Unfortunately, we do not observe how many people pass by each vending machine in our sample in any given week who are considering making a purchase.⁴³ Thus, we estimate market size using data on total

⁴⁰Revenues to manufacturers are calculated as the wholesale cost paid by Mark Vend to the manufacturer. When rebate payments are included, these reduce the amount that Mark Vend pays to each manufacturer by a fixed percentage, which was provided to us by Mark Vend.

⁴¹The exception is that PepsiCo’s revenues increase when Cheetos are removed. This is likely due to the presence of week-to-week variation in relative market shares, market size, and the availability of products in some of the Assorted classes that are not fully controlled for by the matching procedure. However, the result may also partly reflect the fact that PepsiCo owns a very large fraction of the available substitute salty snacks.

⁴²The decline in revenues for Snyder’s, which did not have a focal product in any of the experiments, is due primarily to changes in the set of products that Mark Vend stocked over time. In particular, several Snyder pretzel products were replaced with products manufactured by ‘Minor’ manufacturers and PepsiCo.

⁴³This is more than a simple data limitation, in the sense that even additional data monitoring (e.g., pressure mats or video cameras on the machines) would not provide perfect information on how many consumers may be considering making a purchase as opposed to simply being located near a vending machine.

vends over time at the machine level. Our baseline model specifies market size on the basis of the following regression:

$$y_{rv} = \eta_{rd} + \tau_{rt} + \epsilon_{rv} \quad (4)$$

where y_{rv} is five times total vends at visit v for machine r , η_{rd} is a full set of *machine* \times *day_of_week* fixed effects (six for each machine), and τ_{rt} is a full set of *machine* \times *month* \times *year* fixed effects (38 for each machine).⁴⁴ We specify market size as the predicted value \hat{y}_{rv} from this regression, with two additional restrictions.⁴⁵ We investigated an alternative specification for market size and ran all prediction exercises using the alternative estimate. The results were qualitatively the same as our baseline estimates.⁴⁶

4.2 Nested Logit and Random-coefficients Logit Specifications

We specify two models of demand: nested logit and random-coefficients logit, which are estimated from the full dataset. We consider a model of utility where consumer i receives utility from choosing product j in market t of:

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}. \quad (5)$$

The parameter δ_{jt} is a product-specific intercept that captures the mean utility of product j in market t , and μ_{ijt} captures individual-specific correlation in tastes for products.

Nested Logit

In the case where $(\mu_{ijt} + \varepsilon_{ijt})$ is distributed generalized extreme value, the error terms allow for correlation among products within a pre-specified group, but otherwise assume no correlation. This produces the well-known nested-logit model of McFadden (1978) and Train (2003). In this model consumers first choose a product category l composed of products g_l , and then choose a specific product j within that group. The resulting choice probability for product j in market t is given by the closed-form expression:

$$p_{jt}(\delta, \lambda, a_t) = \frac{e^{\delta_{jt}/\lambda_l} (\sum_{k \in g_l \cap a_t} e^{\delta_{kt}/\lambda_l})^{\lambda_l - 1}}{\sum_{\forall l} (\sum_{k \in g_l \cap a_t} e^{\delta_{kt}/\lambda_l})^{\lambda_l}} \quad (6)$$

where the parameter λ_l governs within-group correlation, and a_t is the set of available products in market t .⁴⁷

⁴⁴The R^2 from this regression is 0.66.

⁴⁵The restrictions are that M must be at least 30 people per visit and must be greater than the total vends for the machine-visit observation.

⁴⁶The alternative model specifies a daily rate of arrival. For each machine-visit observation we denote Δt as the elapsed time since the previous service visit in days and estimate a least-squares regression of $y_{rv}/\Delta t$ on a series of machine fixed effects and a full set of 38 month*year dummies.

⁴⁷Note that this is not the IV regression/“within-group share” presentation of the nested-logit model in Berry (1994), in which σ provides a measure of the correlation of choices within a nest. Roughly speaking,

Random-coefficients Logit

The random-coefficients logit allows for correlation in tastes across observed product characteristics.⁴⁸ This correlation in tastes is captured by allowing the term μ_{ijt} to be distributed according to $f(\mu_{ijt}|\theta)$. A common specification is to allow consumers to have independent normally distributed tastes for product characteristics, so that $\mu_{ijt} = \sum_l \sigma_l \nu_{ilt} x_{jl}$ where $\nu_{ilt} \sim N(0, 1)$ and σ_l represents the standard deviation of the heterogeneous taste for product characteristic x_{jl} . The resulting choice probabilities are a mixture over the logit choice probabilities for many different values of μ_{ijt} , shown here:

$$p_{jt}(\delta, \theta, a_t) = \int \frac{e^{\delta_{jt} + \sum_l \sigma_l \nu_{ilt} x_{jl}}}{1 + \sum_{k \in a_t} e^{\delta_{kt} + \sum_l \sigma_l \nu_{ilt} x_{kl}}} f(v_{ilt}|\theta) \quad (7)$$

Additional Parametrizations

In both the nested-logit and random-coefficient models δ_{jt} consists of product-site intercepts, so that the average taste for an individual product varies from site to site.⁴⁹ For the nested-logit model, we allow for heterogeneous tastes across four major product categories or nests: chocolate, cookie, energy, and salty snack.⁵⁰ For the random-coefficients specification, we allow for three random coefficients, corresponding to consumer tastes for salt, sugar, and fat.⁵¹ An observation for estimation groups machine visits into unique choice sets of the 417 product-site pairs. We report and discuss the estimated parameter values from these models in the section describing the results of the hold-out analyses.

5 Comparing Predicted and Nonparametric Results

Sales are predicted by both models at the machine-visit level using the estimated parameter values and market size y_{rv} . We describe the predicted vends during the treatment and

in the notation used here, $\lambda = 1$ corresponds to the plain logit, and $(1 - \lambda)$ provides a measure of the “correlation” of choices within a nest (as in McFadden (1978)). The parameter λ is sometimes referred to as the “dissimilarity parameter.”

⁴⁸See Berry, Levinsohn, and Pakes (1995).

⁴⁹These correspond to the 417 site-product pairs described in the data section. In addition, one can add a market-level demand shifter ξ_t .

⁵⁰The vending operator defines categories in the same way. “Candy” includes both chocolate and non-chocolate candy items. “Energy” includes products such as peanuts, fruit rolls, crackers, and granola bars.

⁵¹We do not allow for a random coefficient on price because of the relative lack of price variation in the vending machines. We also do not include random coefficients on any discrete variables (such as whether or not a product contains chocolate). As we discuss in Conlon and Mortimer (2009), the lack of variation in a continuous variable (e.g., price) implies that random coefficients on categorical variables may not be identified when product dummies are included in estimation. We did estimate a number of alternative specifications in which we include random coefficients on other continuous variables, such as carbohydrates. In general, the additional parameters were not significantly different from zero, and they had no appreciable effect on the results of any prediction exercises.

matched control weeks and compare them to the actual vends. This exercise does not predict substitution patterns *per se*. Rather, it predicts vends in the environment of the treatment weeks, and again in the environment of the control weeks. Thus, non-experimental changes that occur during the control weeks are included in these predictions. There are two main sources of these non-experimental changes. The first is changes in the characteristics of other products (e.g., a manufacturer may change the fat content of its pretzels, or run a national advertising campaign in a particular week). The second is changes in the availability of other products (e.g., a manufacturer discontinues a particular product, or Mark Vend changes the set of products it carries at a site). The models will adjust for these changes as well as the removal of the focal product. In this sense, neither the nonparametric results nor the predicted vends presented here can be interpreted as the “pure” treatment effect of a product removal.

5.1 Predicted and Actual Sales

Table 7 reports predicted vends from the nested-logit model during treatment and control weeks next to the results of the nonparametric analyses for the Snickers experiment; table 8 repeats the exercise using the random-coefficients logit model.⁵² Predicted vends for the matched control weeks fit well in both tables 7 and 8. While both demand models will necessarily fit average vends of all products quite closely (through the choice of the δ_{jt} parameters), predictions for vends in the matched control weeks need not fit well, because both demand models are estimated off of the full dataset. In this sense, the predicted vends for the matched control weeks provide some insight into the importance of week-to-week variation in market shares in this setting.

Predicted vends fit less well for the treatment weeks than for the matched control weeks. Both models predict more substitution to the outside good than the actual data display: total predicted vends go down by roughly two percent, compared to only 0.1 percent in the actual data. Table 7 shows how the predicted sales patterns in the nested-logit model rely on the nesting structure specified for consumer tastes. Snickers is in the “candy” nest, along with Twix Caramel, Peanut M&Ms, Assorted Chocolate, Assorted Candy, Raisinets, and Skittles. Sales of all of these products are predicted to increase by eight to 17 percent when Snickers is not available. Three of these products (Twix Caramel, Peanut M&Ms, and Assorted Chocolate) are also identified as top substitutes in the nonparametric analyses. Actual vends of Skittles were lower during the treatment weeks than during the matched control weeks. However, it seems unlikely that Snickers is complementary to Skittles in consumption, and the fact that the model does not predict lower sales of Skittles may be viewed as a desirable feature. Assorted Potato Chips are predicted to have relatively high sales during the treatment period. This results from changes in the set of products

⁵²Two full sets of 56 tables (one set for each model) are available in the online appendix. At the individual site level, these tables also report changes in the availability of each product class, as well as any changes to the inclusive value of the assorted product classes that occur between the treatment and control periods due to changes in the individual products that comprise these classes. Further discussion is provided in the online appendix.

included in the Assorted Potato Chip class at two individual sites.⁵³ Two assorted product classes (salty snack and cookie) are predicted by both models to have lower sales during the treatment period compared to the matched control weeks; actual vends are also lower during the treatment period for these classes. This also results from changes in the component products at individual sites: more-popular brands were carried during the matched control weeks than during the treatment weeks.⁵⁴ Relative to the nested-logit model, the random-coefficients model in table 8 predicts smoother sales patterns across product categories. Thus, it successfully predicts more vends of cookie and energy products, which are closer to Snickers in some of the observable product characteristics (i.e., fat and sugar).

5.2 Discussion of Predicted Sales

Both models predict fewer sales of the products identified as ‘top substitutes’ in the matching exercise, and predict more sales to other products. This may be explained partly by noise in the matched data, which we discussed in section 3. However, another explanation is that the parametric models are misspecified. We discuss two potential sources of misspecification.

The first potential source of misspecification is that we have not chosen the correct form for $f(\mu_{ijt}|\theta)$. For the nested-logit model, misspecification is well understood, and occurs when the pre-specified groupings of products that determine substitution patterns do not fully capture consumer behavior. For example, the experiments suggest that many consumers substitute from Snickers to Salted Peanuts, but Salted Peanuts also appear as a substitute for Doritos. Since Doritos and Snickers do not display cross-substitution effects with each other, this creates a dilemma regarding the nest to which Salted Peanuts belong.

There may be several reasons why $f(\mu_{ijt}|\theta)$ is misspecified in the random-coefficients model. The one most strongly suggested by our experimental results is that there are omitted product characteristics for which consumers have heterogeneous tastes. For example, sales of Peanut M&Ms and Salted Peanuts are higher during the Snickers removal than the model suggests, and sales of Salted Peanuts are higher during the Peanut M&M’s removal. However, Peanut M&Ms and Salted Peanuts are not particularly similar along observable dimensions. To formalize this intuition, suppose that the true utility model is given by:

$$\tilde{u}_{ijt} = \delta_{jt} + \sum_l \sigma_l \nu_{ilt} x_{jl} + \gamma_i z_{jt} + \varepsilon_{ijt} \quad \gamma_i \sim N(0, \sigma_z) \quad (8)$$

where z_{jt} is an unobserved characteristic for which consumers have heterogeneous tastes. One can safely exclude the omitted characteristic only when $\sigma_z = 0$. There are two restrictions

⁵³Less-popular brands of chips were stocked during the matched control weeks than during the treatment weeks. The change is most noticeable at sites 5055 and 5655.

⁵⁴The nested-logit model also predicts slightly lower vends during the treatment period for the assorted energy product class, and for one individual product (Grandma’s Chocolate Chip Cookie). The latter prediction is due to the absence of Grandma’s Chocolate Chip Cookie during the treatment period at one site, and the fact that the nested-logit model does not predict much substitution from Snickers to the cookie nest (so predicted vends at other sites do not increase very much). Detailed information on the effects of changes in the component products of the assorted classes, and the availability of all products, is included in the full set of tables in the online appendix.

that the parametric models of u_{ijt} impose. The first is that correlation among tastes is parametrized by μ_{ijt} , which is projected onto the space of observable characteristics x_{jt} .⁵⁵ This approach will always leave some residual correlation, unless the unrestricted $J \times J$ matrix can be estimated. The second restriction is that, conditional on the consumer type μ_{ijt} , all substitution follows a standard logit, including the IIA (Independence of Irrelevant Alternatives) property. The random-coefficients model is a mixture over many different IIA logit models. However, traces of the IIA property remain which may lead to underprediction of substitution to the best substitutes and overprediction of substitution to the worst substitutes.⁵⁶

The second source of difficulty for parametric models is that the set of available products a_t is assumed to be exogenous to the decision-making process. In reality, firms choose product mix carefully in order to maximize profits. The endogeneity of the choice set (or product mix) is something that the researcher cannot generally control, and it is precisely this variation that is used to identify these models. This creates a challenge when we use the models to extrapolate out of sample. For example, Snickers and Twix are both nearly always in stock, and substitution between the two is identified in part by differences in observed substitution to both products from a third product (such as Raisinets) which is sometimes, but not always, stocked. The machines or time periods in which Raisinets are stocked are chosen by Mark Vondt to maximize profits, and may reflect locations or times when the taste for Raisinets is especially high. This leads the model to over-value Raisinets and to mis-specify substitution between Snickers and Twix. The ability to exogenously alter choice sets reduces our reliance on endogenous forms of choice-set variation, and allows us to test the sensitivity of the models to the presence of exogenous variation, which we do in Section 7.

There is a third source of misspecification that is more special to our setting, and that is the absence of price variation.⁵⁷ The ability to identify substitution patterns in the types of demand models estimated here benefits greatly from the presence of price variation, particularly when a strong instrument can be found for addressing the endogeneity of variation in price.⁵⁸

5.3 Predicted Profit Impacts

Differences in predicted vends across products may not have much impact on firms' bottom lines, particularly in a setting with limited price variation. Table 9 reports the predicted profit difference for the downstream firm in the treatment vs. matched control weeks using

⁵⁵McFadden and Train (2000) show that any random utility model can be represented as a mixture of logits as in (5). However, this depends on having a sufficient space of x_{jt} 's.

⁵⁶There are several approaches to dealing with these problems in the literature. One example is the pure characteristics model of Berry and Pakes (2007) which avoids the logit error altogether, but restricts the substitution matrix to vary only according to observable characteristics. Another approach is found in Bajari, Fox, Kim, and Ryan (2010) which proposes a method to non-parametrically recover $f(\mu_{ijt}|\theta)$.

⁵⁷There are, however, many markets with limited price variation, and product availability is often a key consideration in these contexts (e.g., movie theaters, iTunes, and subscription services such as Netflix).

⁵⁸For a discussion of the role of price variation in identifying discrete-choice models of demand, see Ackerman and Rysman (2005), Berry and Haile (2008), and Fox and Gandhi (2010).

the random-coefficients logit model. A comparison of the results to those in table 5 shows that the change in the downstream firm’s average margin is predicted quite well, implying that discrepancies between predicted and actual vends for individual products matters little for the effect on the firm’s average profitability on a per-unit basis. Indeed, the most important determinant for the change in Mark Vend’s average margin is the difference between the wholesale cost of the focal product compared to other goods: Snickers and Peanut M&Ms have very high wholesale costs, and Mark Vend’s average margin increases when those products are removed. This effect is offset somewhat by the rebate payments. The model predicts a more negative overall profit impact for each removal relative to the non-parametric estimates, because total vends during the treatment period are predicted to be lower than the actual data. However, the patterns are correlated with those from the non-parametric analyses: the two most positive/least negative predicted differences correspond to the Peanut M&Ms and Snickers interventions, which were also estimated to have the most positive impacts in the non-parametric results.

Table 10 reports the impact of the product removal on manufacturer revenue predicted by the random-coefficients model, with and without rebate payments. Relative to the non-parametric estimates in table 6, some experiments have predicted impacts that are quite similar (e.g., the cookie experiments), while others predict a larger negative impact for the manufacturer of the focal product and corresponding larger positive impacts for competing manufacturers. This reflects the fact that many of the non-focal products with under-predicted sales are owned by the manufacturer of the focal product (e.g., both Snickers and Peanut M&Ms are owned by Mars Inc.). As expected, the impact of each product removal is less negative for manufacturers after accounting for rebate payments.

6 Simulating Stockout Events

The exercise in which we predict vends for the treatment and control weeks is an exercise in which we are rarely interested from an economic point of view. Policy makers are generally interested in understanding the impact of a control variable (in this case, the decision to stock a product), rather than understanding the joint effect of a control variable and other factors. Both the non-parametric approach and the prediction exercise above conflates these things, modeling the product removal simultaneously with changes in other factors that shift consumer tastes (such as national manufacturer advertising campaigns). For that reason, the prediction exercise is also a strenuous test on which to expect the models to perform well, because of the need to adjust for multiple factors simultaneously. In this section, we simulate the effect of a product removal directly, using the estimated model parameters and holding all other factors fixed.

Both the nested logit and random-coefficients logit specifications have straightforward predictions regarding how demand should respond to a product removal. The expected change in sales of product j when product k is removed is:

$$\Delta q_{jt}^{(k)} = M_t \left(p_{jt}(\delta, \theta, a_{t \setminus k}) - p_{jt}(\delta, \theta, a_t) \right). \quad (9)$$

We calculate $\Delta q_{jt}^{(k)}$ for the set of products available during the treatment weeks at each site. Table 11 reports the results of this calculation for the Snickers removal, summing over all sites.⁵⁹ We choose M_t so that the sales levels of the focal product(s) match the sales levels in the actual data during the matched control weeks. The left panel reports the change in sales for a simulated removal of Snickers based on the nested logit and random-coefficients logit models. Under the random-coefficients specification, Snickers has sales of 323 units when it is available. Of these 323 “Snickers consumers,” 73 leave the machine and 250 purchase another product (e.g., 34 purchase an Assorted Energy product, 33 purchase Peanut M&Ms, etc.). The downstream firm loses \$69 in variable profit from the sales of Snickers, but gains \$95 in variable profit across all substitute goods. Thus, in spite of the fact that 73 Snickers consumers leave the machine, the removal of Snickers is profitable for the downstream firm.

Table 12 summarizes the overall profit impacts of simulations in which we remove each of the focal products. As expected, vends decrease in every case because of substitution to the outside good. As in the nonparametric analyses and the prediction exercise, average margins increase in the three candy removals as consumers substitute to products with lower wholesale costs for the downstream firm, although this effect is ameliorated by the rebate payments. Mark Vend’s variable profits are still forecast to increase when Snickers is removed, although the rebate payments provide an incentive to carry Peanut M&Ms. All other product removals result in lower profits for Mark Vend, even in the absence of rebate payments. The revenue impacts for upstream firms are reported in table 13. The manufacturer of each focal product loses revenue when the product is removed, while other manufacturers gain. The gains to other manufacturers are spread out fairly widely.

7 Hold-out Analyses

In most applications, the type of exogenous variation in product availability that we induce through our experiments does not exist in the data. Demand models are thus identified entirely from naturally-occurring variation in choice sets. This variation may arise from many sources (e.g., changes in stocking decisions, price, and other characteristics), but is often endogenously determined by firms. Our setting permits us to examine the sensitivity of our demand estimates to the presence of exogenous variation in choice sets.

Table 14 reports the estimated coefficients that govern correlation in consumer tastes when we estimate our demand models on subsets of the data that exclude different sets of our exogenous product removals. The random-coefficients model is shown in the top panel, and the nested logit model in the second panel. The bottom panel reports the number of products and choice sets for each sample of the data. The parameter values from the base model, in which the full set of data is used for estimation, are reported in the first column.⁶⁰ The second column reports the estimated parameters when the model is estimated on only control observations (i.e., we exclude all experimental periods). The estimates of σ_{sugar} and

⁵⁹The full set of 56 tables is provided in the online appendix.

⁶⁰The log likelihood values are reported in the last row of each panel. They are the same for both models, but are not comparable across the different hold-out sub-samples because they apply to a different number of observations.

λ_{candy} change significantly (from 5.25 to 2.91 for σ_{sugar} and from 0.57 to 0.66 for λ_{candy}). The other parameters are more robust to the exclusion of the treatment observations. The third column of table 14 reports the estimated parameters when the models are estimated without the benefit of the data from the candy experiments. We see the same effect on the parameter values: the estimated correlation in tastes for sugar and the candy nest are lower (i.e., σ_{sugar} is lower and λ_{candy} is higher).⁶¹ The remaining columns repeat the exercise by withholding data from the salty snack and cookie experiments in turn. The estimated parameter values are much more robust to these hold-out exercises.

The bottom panel provides some guidance for assessing these results. The total number of products is reported next to each category name (e.g., there are 417 products total, 115 of which are candy products). The experimental interventions increase the total number of unique choice sets from 1096 to 1734. Among candy products alone, there are 427 unique choice sets in the full dataset, but only 262 in the subset of the data that includes only control observations. The salty snack category has about 60 percent more products than the candy category (187 vs. 115), but 120 percent more unique within-category choice sets when using only the control observations (578 vs. 262). Removing data from the salty snack interventions does reduce the number of within-category choice sets, but not as dramatically (578 vs. 794).

The choice-set data help to better understand the challenges of identification in demand models when there is relatively little variation in stocking (or other product characteristics) within a group of closely-related products. In our setting, the experimentally stocked-out candy bars have large market shares relative to their salty-snack counterparts, and are nearly always available simultaneously to consumers during the control observations. Furthermore, there is limited variation (relative to the number of products) in consumer choice sets generated by other candy products. Thus, identifying correlation in tastes for the characteristics that are important to these goods is greatly aided by our ability to exogenously create choice-set variation.⁶²

8 Conclusion

Our analysis of product availability makes use of a field experiment in which top-selling products are removed from a set of 60 vending machines. We analyze changes in purchasing patterns and profits in two ways: first, using nonparametric techniques common to the treatment effects and experiments literature; second, using structural demand estimation. We find substitution patterns that seem quite sensible, and we note that in the short-run, relatively few people leave the market when a product is out of stock. We find evidence of the incentive problems facing firms in vertically-separated markets: some product removals result in lower revenues for upstream manufacturers but higher profits for the downstream vending

⁶¹Recall that the λ parameters give an indication of the ‘dissimilarity’ of products, so that a higher estimate of λ indicates less substitution between two products within a nest.

⁶²As discussed earlier, price variation will be helpful in many settings for generating additional variation in the choice sets facing consumers.

operator in the absence of rebate payments because consumers substitute to products with lower wholesale costs.

We discuss the trade-offs of both empirical approaches. The treatment-effects approach places no parametric restrictions on substitution patterns, but noisy estimates occasionally imply nonsensical effects of a product’s removal. The structural approach avoids this problem, but at the cost of imposing restrictions on the substitution patterns and profit impacts associated with a product’s removal. Using hold-out analyses, we find that the structural model is sensitive to the absence of the experimental data in parts of product space where relatively little non-experimental variation occurs in consumers’ choice sets.

Product availability is a critical feature of many markets; firms make both long-run and short-run decisions about the capacity or inventory of different products to stock, which brands to carry, and how to respond to changes in the product availability of rivals. Product availability can vary over time due to mergers, foreclosure, or other factors. Despite the prevalence of these issues, relatively little empirical evidence exists in the IO literature on the importance of product stocking decisions for firm profits, consumer choices, or vertical relationships. Using vending machines, we are able to exogenously alter a firm’s product mix in order to shed light on two central outcomes of interest (i.e., substitution patterns and profit impacts for both upstream and downstream firms), which apply broadly to more complex settings. The experimental approach also provides an opportunity to compare the trade-offs and assess the sources of identification of different empirical methods.

References

- ABADIE, A., D. DRUKKER, J. HERR, AND G. IMBENS (2004): “Implementing Matching Estimators for Average Treatment Effects in Stata,” *The Stata Journal*, 4(3), 290–311.
- ABADIE, A., AND G. IMBENS (2006): “Large Sample Properties of Matching Estimators for Average Treatment Effects,” *Econometrica*, 74(1), 235–267.
- ACKERBERG, D., AND M. RYSMAN (2005): “Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects,” *RAND Journal of Economics*, 36(4).
- AGUIRREGABIRIA, V. (1999): “The Dynamics of Markups and Inventories in Retailing Firms,” *Review of Economic Studies*, 66, 278–308.
- ANDERSON, E. T., G. J. FITZSIMONS, AND D. SIMESTER (2006): “Measuring and Mitigating the Costs of Stockouts,” *Management Science*, 52(11), 1751–1763.
- 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.
- ANUPINDI, R., M. DADA, AND S. GUPTA (1998): “Estimation of Consumer Demand with Stock-Out Based Substitution: An Application to Vending Machine Products,” *Marketing Science*, 17(4), 406–423.
- ARROW, K., T. HARRIS, AND J. MARSCHAK (1951): “Optimal Inventory Policy,” *Econometrica*, 19(3), 250–272.
- BAJARI, P., J. FOX, K. I. KIM, AND S. RYAN (2010): “A Simple Nonparametric Estimator for the Distribution of Random Coefficients,” Working Paper.
- BERRY, S. (1994): “Estimating discrete-choice models of product differentiation,” *RAND Journal of Economics*, 25(2), 242–261.
- BERRY, S., AND P. HAILE (2008): “Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers,” Working Paper.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841–890.
- BERRY, S., AND A. PAKES (2007): “The Pure Characteristics Demand Model,” *International Economic Review*, 48(4), 1193–1125.
- BERTRAND, M., D. KARLAN, S. MULLAINATHAN, E. SHAFIR, AND J. ZINMAN (2010): “What’s Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment*,” *Quarterly Journal of Economics*, 125(1), 263–305.

- BROWNELL, K., T. FARLEY, W. WILLETT, B. POPKIN, F. CHALOUPKA, J. THOMPSON, AND D. LUDWIG (2009): “The public health and economic benefits of taxing sugar-sweetened beverages,” *New England Journal of Medicine*, 361, 1599–1605.
- BROWNELL, K., AND T. FRIEDEN (2009): “Ounces of prevention: The public policy case for taxes on sugared beverages,” *New England Journal of Medicine*, 360, 1805–1808.
- BRUNO, H. A., AND N. VILCASSIM (2008): “Structural Demand Estimation with Varying Product Availability,” *Marketing Science*, 27(6), 1126–1131.
- CAI, H., H. FANG, AND C. YUYU (2009): “Observational Learning: Evidence from a Randomized Natural Field Experiment,” *American Economic Review*, 99(3), 864–82.
- CONLON, C., AND J. H. MORTIMER (2009): “Demand Estimation Under Incomplete Product Availability,” Working Paper.
- DEHEJIA, R. H., AND S. WAHBA (1999): “Causal Effects in Nonexperimental Studies: Re-evaluation of the Evaluation of Training Programs,” *Journal of the American Statistical Association*, 94, 1053–1062.
- EDGEWORTH, F. (1888): “The Mathematical Theory of Banking,” *Journal of the Royal Statistical Society*, 53, 113–127.
- EINAV, L., AND J. LEVIN (2010): “Empirical Industrial Organization: A Progress Report,” *The Journal of Economic Perspectives*, 24(2), 145–162.
- ENGBER (2009): “Let Them Drink Water! What a fat tax really means for America.,” <http://www.slate.com/id/2228713/>.
- ENGELBRECHT-WIGGANS, R., J. A. LIST, AND D. H. REILEY (2005): “Demand Reduction in Multi-Unit Auctions: Evidence from a Sportscard Field Experiment: Reply,” *American Economic Review*, 95(1), 472–476.
- ERIV, I., AND A. ROTH (1998): “Predicting how People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria,” *American Economic Review*, 88(4), 848–881.
- FITZSIMONS, G. J. (2000): “Consumer Response to Stockouts,” *Journal of Consumer Research*, 27(2), 249–266.
- FOX, J., AND A. GANDHI (2010): “Nonparametric Identification and Estimation of Random Coefficients in Nonlinear Economic Models,” 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.
- IMBENS, G. (2004): “Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review,” *Review of Economics and Statistics*, 86(1), 4–29.

- KARLAN, D., AND J. A. LIST (2007): “Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment,” *American Economic Review*, 97(5), 1774–1793.
- KEANE, M. (2010): “A Structural Perspective on the Experimentalist School,” *The Journal of Economic Perspectives*, 24(2), 47–58.
- KRAISELBURD, S., V. NARAYANAN, AND A. RAMAN (2004): “Contracting in a Supply Chain with Stochastic Demand and Substitute Products,” *Production and Operations Management*, 13(1), 46–62.
- LALONDE, R. (1986): “Evaluating the Econometric Evaluations of Training Programs,” *American Economic Review*, 76(4), 604–620.
- 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.
- LEVITT, S. D., AND J. A. LIST (2009): “Field experiments in economics: The past, the present, and the future,” *European Economic Review*, 53(1), 1–18.
- MATSA, D. (2010): “Competition and Product Quality in the Supermarket Industry,” Working Paper.
- McFADDEN, D. (1978): “Modelling the Choice of Residential Location,” in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlqvist, L. Lundsqvist, F. Snickars, and J. Weibull. North-Holland.
- McFADDEN, D., AND K. TRAIN (2000): “Mixed MNL Models for Discrete Response,” *Journal of Applied Econometrics*, 15, 447–470.
- MUSALEM, A., M. OLIVARES, E. BRADLOW, C. TERWIESCH, AND D. CORSTEN (2010): “Structural Estimation of the Effect of Out-of-Stocks,” *Management Science*, 52(7), 1180–1197.
- 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.
- OSTROVSKY, M., AND M. SCHWARZ (2009): “Reserve Prices in Internet Advertising Auctions: A Field Experiment,” Working Paper.
- SCHWEITZER, M., AND G. CACHON (2000): “Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence,” *Management Science*, 46(3), 404–420.

- SIMS, C. (2010): “But Economics Is Not an Experimental Science,” *The Journal of Economic Perspectives*, 24(2), 59–68.
- SMITH, S. (2010): “City may curb sales of sugary beverages,” *The Boston Globe*, September 20th.
- SMITH, V. (1962): “Loss Aversion in Riskless Choice: A Reference-Dependent Model,” *Journal of Political Economy*, 70(2), 111–137.
- SPENGLER, J. (1950): “Vertical Integration and Antitrust Policy,” *Journal of Political Economy*, 58, 347–352.
- STOCK, J. (2010): “The Other Transformation in Econometric Practice: Robust Tools for Inference,” *The Journal of Economic Perspectives*, 24(2), 83–94.
- TRAIN, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.
- TVERSKY, A., AND D. KAHNEMAN (1991): “Loss Aversion in Riskless Choice: A Reference-Dependent Model,” *Quarterly Journal of Economics*, 106(4), 1039–1061.

Table 1: Changes in Purchasing Patterns for Snickers Removal, All Sites

Product	Control Mean	Control %ile	Treatment Mean	Treatment %ile	Mean Difference	% Δ
<i>Vends</i>						
Peanut M&Ms	359.9	73.6	478.3*	99.4	118.4*	32.9
Twix Caramel	187.6	55.3	297.1*	100.0	109.5*	58.4
Assorted Chocolate	334.8	66.7	398.0*	95.0	63.2*	18.9
Assorted Energy	571.9	63.5	616.2	76.7	44.3	7.8
Zoo Animal Cracker	209.1	78.6	243.7*	98.1	34.6*	16.5
Salted Peanuts	187.9	70.4	216.3*	93.7	28.4	15.1
Choc Chip Famous Amos	171.6	71.7	193.1*	95.0	21.5*	12.5
Ruger Vanilla Wafer	107.3	59.7	127.9	78.6	20.6*	19.1
Assorted Candy	215.8	43.4	229.6	60.4	13.7	6.4
Assorted Potato Chips	279.6	64.2	292.4*	66.7	12.8	4.6
Assorted Pretzels	548.3	87.4	557.7*	88.7	9.4	1.7
Raisinets	133.3	66.0	139.4	74.2	6.1	4.6
Cheetos	262.2	60.1	260.5	58.2	-1.8	-0.7
Grandmas Choc Chip	77.9	51.3	72.5	37.8	-5.4	-7.0
Doritos	215.4	54.1	203.1	39.6	-12.3*	-5.7
Assorted Cookie	180.3	61.0	162.4	48.4	-17.9	-10.0
Skittles	100.1	62.9	75.1*	30.2	-25.1*	-25.0
Assorted Salty Snack	1382.8	56.0	1276.2*	23.3	-106.7*	-7.7
Snickers	323.4	50.3	2.0*	1.3	-321.4*	-99.4
Total	5849.6	74.2	5841.3	73.0	-8.3	-0.1
<i>Shares</i>						
Peanut M&Ms	6.2	60.4	8.2*	100.0	2.0*	33.1
Twix Caramel	3.2	41.5	5.1*	96.9	1.9*	58.6
Assorted Chocolate	5.7	53.5	6.8*	96.9	1.1*	19.0
Assorted Energy	9.8	49.1	10.5	75.5	0.8	7.9
Zoo Animal Cracker	3.6	65.4	4.2*	98.1	0.6*	16.7
Salted Peanuts	3.2	56.6	3.7*	89.9	0.5	15.3
Choc Chip Famous Amos	2.9	64.8	3.3*	90.6	0.4*	12.7
Ruger Vanilla Wafer	1.8	51.6	2.2	74.8	0.4*	19.3
Assorted Candy	3.7	35.8	3.9	49.7	0.2	6.5
Assorted Potato Chips	4.8	62.9	5.0*	63.5	0.2	4.7
Assorted Pretzels	9.4	75.5	9.5*	82.4	0.2	1.9
Raisinets	2.3	58.5	2.4	67.9	0.1	4.7
Cheetos	4.5	38.0	4.5	36.1	-0.0	-0.5
Grandmas Choc Chip	1.3	39.7	1.2	23.1	-0.1	-6.8
Doritos	3.7	37.7	3.5	20.8	-0.2*	-5.6
Assorted Cookie	3.1	55.3	2.8	44.0	-0.3	-9.8
Skittles	1.7	56.6	1.3*	20.1	-0.4*	-24.9
Assorted Salty Snack	23.6	32.7	21.8*	4.4	-1.8*	-7.6
Snickers	5.5	44.0	0.0*	0.6	-5.5*	-99.4
Total	100.0	50.3	100.0	50.3	0.0	0.0

Notes: Control weeks are selected through nearest-neighbor matching using four control observations for each treatment week. Percentiles are relative to the full distribution of control weeks.

Outcomes outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 2: Top 5 Substitutes (Based on Change in Sales), All Removals

<i>Candy Experiments</i>			
Focal Product:	Snickers	Peanut M&Ms	Snickers + Peanut M&Ms
Top 5 Substitutes:	Peanut M&Ms* Twix Caramel* Ass. Chocolate* Ass. Energy Zoo Animal Cracker*	Snickers* Plain M&Ms* Twix Caramel* Salted Peanuts* Raisinets*	Twix Caramel* Plain M&Ms* Ass. Chocolate* Raisinets* Ass. Cookie*
Effects:	Focal (-321.4) Top 5 (+370.0) Total (-8.3)	Focal (-317.7) Top 5 (+265.2) Total (+54.1)	Focals (-287.6,-305.9) Top 5 (+305.5) Total (-218.4)
<i>Salty Snack Experiments</i>			
Focal Product:	Doritos	Cheetos	Doritos + Cheetos
Top 5 Substitutes:	Ass. Potato Chips* Peanut M&Ms* Salted Peanuts* Sun Chips* Ass. Energy	Sun Chips* Ass. Potato Chips* Ass. Energy Frito* Salted Peanuts*	Ass. Salty Snack* Ass. Energy Snickers* Cheez-It Original* Hot Stuff Jays*
Effects:	Focal (-210.6) Top 5 (+183.6) Total (-75.1)	Focal (-273.2) Top 5 (+284.6) Total (-122.5)	Focals (-172.7,-215.5) Top 5 (+324.1) Total (-35.9)
<i>Cookie Experiments</i>			
Focal Product:	Zoo Animal Crackers	Choc Chip Famous Amos	
Top 5 Substitutes:	Peanut M&M* Ass. Energy Snickers* Twix Caramel* Ruger Vanilla Wafer*	Sun Chips* Salted Peanuts* Ass. Potato Chips Raspberry Knotts* Grandma Choc Chip*	
Effects:	Focal (-210.6) Top 5 (+216.8) Total (-28.7)	Focal (-141.6) Top 5 (+155.7) Total (-179.5)	

Notes: Control weeks are selected through nearest-neighbor matching using four control observations for each treatment week. Effects report the change in average vends of the treatment and control weeks for: the focal product, the Top 5 substitutes based on changes in sales levels, and total vends. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 3: Top 5 Substitutes (Based on Percentage Change in Own Sales), All Removals

<i>Candy Experiments</i>			
Focal Product:	Snickers	Peanut M&Ms	Snickers + Peanut M&Ms
Top 5 Substitutes:	Twix Caramel* Peanut M&Ms* Ruger Vanilla Wafer* Ass. Chocolate* Zoo Animal Crackers*	Plain M&Ms* Snickers* Raisinets* Farleys Mixed Fruit* Salted Peanuts*	Plain M&Ms* Twix Caramel* Ass. Chocolate* Raisinets* Reeses Peanut Butter Cups*
Effects:	Focal (-5.5%) Top 5 (+6.0%) Total (-0.1%)	Focal (-5.9%) Top 5 (+4.5%) Total (+1.0%)	Focal (-5.4%,-5.7%) Top 5 (+6.4%) Total (-4.1%)
<i>Salty Snack Experiments</i>			
Focal Products:	Doritos	Cheetos	Doritos + Cheetos
Top 5 Substitutes:	Salted Peanuts* Ass. Potato Chips* Sun Chips* Peanut M&Ms* Choc Chip F A	Sun Chips* Frito* Salted Peanuts* Ass. Potato Chips* Doritos*	Hot Stuff Jays* Cheez-It Original* Frito* Ass. Salty Snack* Smartfood
Effects:	Focal (-3.6%) Top 5 (+3.3%) Total (-1.2%)	Focal (-4.6%) Top 5 (+4.8%) Total (-2.1%)	Focal (-3.5%,-4.4%) Top 5 (+5.9%) Total (-0.7%)
<i>Cookie Experiments</i>			
Focal Products:	Zoo Animal Crackers	Choc Chip Famous Amos	
Top 5 Substitutes:	Ruger Vanilla Wafer* Raspberry Knotts Twix Caramel* Peanut M&Ms* Snickers*	Sun Chips* Salted Peanuts* Raspberry Knotts* Grandma Choc Chip* Ass. Potato Chips	
Effects:	Focal (-3.4%) Top 5 (+3.0%) Total (-0.5%)	Focal (-2.5%) Top 5 (+3.2%) Total (-3.2%)	

Notes: Control weeks are selected through nearest-neighbor matching using four control observations for each treatment week. Effects report the change in average vends of the treatment and control weeks for: the focal product, the Top 5 substitutes based on percentage changes in sales levels, and total vends. Percentage changes for Focal and Top 5 products refer to changes in their average inside market share. Percentage change for Total reports the percentage change in the number of total vends. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 4: Impact of Snickers Removal on Downstream Firm Profit, All Sites

Product	Manufacturer	Retail Margin	Difference in Vends	Difference in Profit
Peanut M&Ms	Mars	0.21	118.4*	25.2*
Twix Caramel	Mars	0.21	109.5*	23.3*
Assorted Chocolate	Hershey's / Mars / Other	0.25	63.2*	13.2*
Assorted Energy	Various	0.48	44.3	17.6
Zoo Animal Cracker	Kellogg's	0.52	34.6*	17.6*
Salted Peanuts	Kraft	0.49	28.4	13.2*
Choc Chip Famous Amos	Kellogg's	0.53	21.5*	12.7*
Ruger Vanilla Wafer	Other	0.50	20.6*	10.5
Assorted Candy	Various	0.41	13.7	5.9
Assorted Potato Chips	PepsiCo	0.35	12.8	4.7
Assorted Pretzels	Snyder's / PepsiCo	0.38	9.4	1.6
Raisinets	Other	0.31	6.1	1.9
Cheetos	PepsiCo	0.41	-1.8	-0.7
Grandmas Choc Chip	PepsiCo	0.59	-5.4	-3.4
Doritos	PepsiCo	0.41	-12.3*	-5.0*
Assorted Cookie	Various	0.48	-17.9	-10.3
Skittles	Mars	0.21	-25.1*	-5.2*
Assorted Salty Snack	Various	0.41	-106.7*	-51.6*
Snickers	Mars	0.21	-321.4*	-68.4*
Total		0.38	-8.3	2.6

Notes: Retail margin is reported in cents per unit. Difference in vends is reported in units per week, and difference in revenue is reported in dollars per week. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 5: Impact of All Product Removals on Total Downstream Firm Profit

Product(s) Removed:	Difference in:	Before Rebate		After Rebate	
	Vends	Margin	Profit	Margin	Profit
Snickers	-8.27	0.02	2.65	-0.05	-2.50
Animal Crackers	-28.68	-0.80	-60.40	-0.64	51.20
Doritos	-75.08	-0.48	-53.64	-0.40	-51.44
Peanut M&Ms	54.13	0.47	43.78	0.35	39.79
Cheetos	-122.49	-0.16	-55.06	-0.07	-53.95
Choc Chip Famous Amos	-179.47	-0.34	-89.21	-0.28	-91.40
Cheetos + Doritos	-35.93	-0.21	-12.76	-0.15	-12.68
Snickers + Peanut M&Ms	-218.37	1.21	-15.00	0.89	-42.14

Notes: Difference in margin is reported in cents per unit. Difference in vends is reported in units per week, and difference in profits is reported in dollars per week. No profit difference is outside the 90% Confidence Interval described in section 3.1.1.

Table 6: Impact on Manufacturer Revenue

Product(s) Removed:	Before Rebate						
	Hershey's	Kellogg	Kraft	Mars	PepsiCo	Snyder's	Minor
Snickers	-5.85	22.37	11.42	-28.27	19.31	-28.11	25.94
Animal Crackers	-3.93	-34.18	10.89	81.37	33.45	-23.60	2.60
Doritos	-2.65	22.01	15.27	41.37	-2.27	-33.52	2.66
Peanut M&Ms	0.65	9.63	10.42	-49.11	56.98	-5.74	23.74
Cheetos	0.84	4.03	15.99	13.66	21.41	-32.16	-37.92
Choc Chip Famous Amos	-9.55	-51.41	19.10	2.96	65.62	-34.80	-35.47
Cheetos + Doritos	12.60	14.63	-1.43	12.48	-49.71	3.83	-2.26
Snickers + Peanut M&Ms	45.93	18.19	5.78	-218.43	0.99	-9.57	49.09
Product(s) Removed:	After Rebate						
	Hershey's	Kellogg	Kraft	Mars	PepsiCo	Snyder's	Minor
Snickers	-4.99	20.70	10.37	-23.82	17.53	-25.75	26.40
Animal Crackers	-3.42	-31.33	10.07	68.24	30.24	-21.71	3.87
Doritos	-2.30	20.46	13.90	34.64	-1.52	-30.66	3.60
Peanut M&Ms	0.29	8.95	9.64	-41.23	51.39	-5.37	25.10
Cheetos	0.51	3.70	14.52	11.36	19.41	-29.53	-35.62
Choc Chip Famous Amos	-7.99	-47.31	17.16	2.37	59.00	-31.89	-32.90
Cheetos + Doritos	10.37	13.49	-1.44	10.45	-44.84	3.45	-2.39
Snickers + Peanut M&Ms	37.92	16.68	5.59	-183.38	0.77	-8.70	47.94

Notes: Revenues to manufacturer are calculated as the wholesale cost paid by Mark Vend to the manufacturer, not including any potential rebate payments. Revenue impacts for the manufacturer of the focal product of any given experiment are shown in bold typeface (e.g., Mars Inc. manufactures Snickers). Minor manufacturers include all 17 manufacturers listed in table 15.

Table 7: Predicted Sales During the Snickers Treatment and Matched Control Weeks (All Sites), Nested Logit Model

	Results of Matching (Means)			Fitted Values (Means)		
	Treatment	Control	% Δ	Treatment	Control	% Δ
Twix Caramel	297.1*	187.6	58.4	208.6*	190.7	9.4
Peanut M&Ms	478.3*	359.9	32.9	388.6*	349.1	11.3
Ruger Vanilla Wafer	127.9	107.3	19.1	114.3	111.5	2.5
Assorted Chocolate	398.0*	334.8	18.9	371.0*	328.6	12.9
Zoo Animal Cracker	243.7*	209.1	16.5	206.9	205.2	0.8
Salted Peanuts	216.3*	187.9	15.1	190.7*	184.8	3.2
Choc Chip Famous Amos	193.1*	171.6	12.5	171.4	167.5	2.4
Assorted Energy	616.2	571.9	7.8	582.7	584.3	-0.3
Assorted Candy	229.6	215.8	6.4	253.1*	217.0	16.6
Raisinets	139.4	133.3	4.6	142.8	129.5	10.3
Assorted Potato Chips	292.4*	279.6	4.6	348.4*	290.9	19.8
Assorted Pretzels	557.7*	548.3	1.7	533.2*	522.4	2.1
Cheetos	260.5	262.2	-0.7	271.8	270.1	0.6
Doritos	203.1	215.4	-5.7	229.3	226.5	1.3
Grandmas Choc Chip	72.5	77.9	-7.0	83.7	83.9	-0.3
Assorted Salty Snack	1276.2*	1382.8	-7.7	1356.3	1410.5	-3.8
Assorted Cookie	162.4	180.3	-10.0	162.5	178.4	-8.9
Skittles	75.1*	100.1	-25.0	100.3	93.2	7.6
Snickers	2.0*	323.4	-99.4	4.4*	336.2	-98.7
Total	5841.3	5849.6	-0.1	5719.9	5880.2	-2.7

Notes: Products are sorted by percentage change in actual values. Predicted results report sales during the treatment weeks and control weeks using the parameters estimated in the baseline nested-logit model. Just as for the actual data, predicted vends of the focal product result from the assignment of service visits to weeks. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 8: Predicted Sales During the Snickers Treatment and Matched Control Weeks (All Sites), Random Coefficients Model

	Results of Matching (Means)			Fitted Values (Means)		
	Treatment	Control	% Δ	Treatment	Control	% Δ
Twix Caramel	297.1*	187.6	58.4	206.2*	189.8	8.7
Peanut M&Ms	478.3*	359.9	32.9	382.4*	347.9	9.9
Ruger Vanilla Wafer	127.9	107.3	19.1	120.7	110.2	9.5
Assorted Chocolate	398.0*	334.8	18.9	367.7*	327.6	12.2
Zoo Animal Cracker	243.7*	209.1	16.5	211.8*	204.3	3.7
Salted Peanuts	216.3*	187.9	15.1	193.8*	183.9	5.4
Choc Chip Famous Amos	193.1*	171.6	12.5	178.9*	165.9	7.9
Assorted Energy	616.2	571.9	7.8	604.6	579.6	4.3
Assorted Candy	229.6	215.8	6.4	246.0*	217.9	12.9
Raisinets	139.4	133.3	4.6	138.7	129.4	7.2
Assorted Potato Chips	292.4*	279.6	4.6	349.6*	289.8	20.6
Assorted Pretzels	557.7*	548.3	1.7	525.7*	519.4	1.2
Cheetos	260.5	262.2	-0.7	264.6	266.4	-0.7
Doritos	203.1	215.4	-5.7	229.6	225.5	1.8
Grandmas Choc Chip	72.5	77.9	-7.0	90.5*	83.2	8.7
Assorted Salty Snack	1276.2*	1382.8	-7.7	1353.1	1403.2	-3.6
Assorted Cookie	162.4	180.3	-10.0	169.1	178.5	-5.2
Skittles	75.1*	100.1	-25.0	97.7	92.9	5.1
Snickers	2.0*	323.4	-99.4	4.4*	334.1	-98.7
Total	5841.3	5849.6	-0.1	5735.1	5849.6	-2.0

Notes: Products are sorted by percentage change in actual values. Predicted results report sales during the treatment weeks and control weeks using the parameters estimated in the baseline random-coefficients logit model. Just as for the actual data, predicted vends of the focal product result from the assignment of service visits to weeks. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *’s.

Table 9: Impact of All Product Removals on Total Downstream Firm Profit (Predicted, Random Coefficients)

	Difference in:	Before Rebate		After Rebate	
Product(s) Removed:	Vends	Margin	Profit	Margin	Profit
Snickers	-114.46	0.45	-15.82	0.31	-30.43
Animal Crackers	-152.71	-0.85	-109.71	-0.67	-104.85
Doritos	-193.68	-0.48	-98.08	-0.36	-99.34
Peanut M&M's	-82.42	0.69	1.70	0.50	-9.93
Cheetos	-205.94	-0.19	-94.33	-0.19	-96.85
Choc Chip Famous Amos	-243.73	-0.52	-130.73	-0.53	-133.28
Cheetos + Doritos	-187.34	-0.06	-74.43	-0.11	-79.75
Snickers + Peanut M&M's	-369.02	1.98	-33.76	1.58	-73.94

Notes: Difference in margin is reported in cents per unit. Difference in vends is reported in units per week, and difference in profits is reported in dollars per week. No profit difference is outside the 90% Confidence Interval described in section 3.1.1.

Table 10: Impact on Manufacturer Revenue (Predicted, Random Coefficients)

Product(s)	Before Rebate						
Removed:	Hershey's	Kellogg	Kraft	Mars Inc.	PepsiCo	Snyder's	Minor
Snickers	-0.18	10.46	9.33	-109.08	41.52	-19.70	2.54
Animal Crackers	-1.19	-49.44	6.15	43.09	44.73	-19.02	-27.52
Doritos	-6.39	3.74	13.95	31.28	-16.68	-31.36	-26.42
Peanut M&Ms	3.95	9.95	9.48	-96.56	49.02	-18.13	12.46
Cheetos	-6.83	-0.65	17.13	17.29	-10.67	-31.69	-36.35
Choc Chip Famous Amos	-9.52	-48.99	6.01	30.75	13.51	-30.69	-17.96
Cheetos + Doritos	11.60	1.69	6.46	2.01	-81.60	2.24	-2.67
Snickers + Peanut M&Ms	26.45	16.53	8.22	-291.40	19.39	-10.53	29.21
	After Rebate						
	Hershey's	Kellogg	Kraft	Mars	PepsiCo	Snyder's	Minor
Snickers	-0.32	9.64	8.39	-91.68	37.40	-18.06	3.70
Animal Crackers	-1.22	-45.49	5.54	36.14	40.25	-17.50	-25.79
Doritos	-5.37	3.47	12.53	26.25	-14.82	-28.66	-24.50
Peanut M&Ms	3.01	9.21	8.64	-81.17	44.20	-16.64	13.30
Cheetos	-5.77	-0.61	15.37	14.38	-9.55	-29.11	-33.62
Choc Chip Famous Amos	-7.96	-45.08	5.36	25.73	12.13	-28.12	-15.65
Cheetos + Doritos	9.49	1.58	5.90	1.66	-73.48	2.00	-2.45
Snickers + Peanut M&Ms	21.71	15.26	7.68	-244.73	17.42	-9.60	29.10

Notes: Revenues to manufacturer are calculated as the wholesale cost paid by Mark Vend to the manufacturer, not including any potential rebate payments. Revenue impacts for the manufacturer of the focal product of any given experiment are shown in bold typeface (e.g., Mars Inc. manufactures Snickers). Minor manufacturers include all 17 manufacturers listed in table 15.

Table 11: Simulated Effects of Removing Snickers (All Sites)

	Change in Sales		Change in Profit	
	Nested Logit	RC Logit	Nested Logit	RC Logit
Peanut M&Ms	39.35*	32.87*	8.37*	6.99*
Assorted Chocolate	34.96*	29.73	10.14	8.47
Assorted Candy	24.20	16.16	10.45*	6.96
Twix Caramel	16.47	16.38	3.50	3.48
Raisinets	10.78	8.14	3.34	2.52
Skittles	10.29	10.51	1.89	1.98
Assorted Salty Snack	10.02	27.67	4.11	11.31
Assorted Energy	4.52	34.35	2.20	16.57
Assorted Pretzels	3.29	7.61	1.24	2.88
Cheetos	1.85	4.80	0.76	1.97
Doritos	1.53	4.65	0.63	1.91
Assorted Potato Chips	1.43	3.70	0.53	1.37
Assorted Cookie	1.41	12.31	0.65	5.59
Zoo Animal Cracker	1.36	8.89	0.69	4.43
Choc Chip Famous Amos	1.12	12.13	0.61	6.53
Grandmas Choc Chip	0.67	12.00	0.39	7.02
Ruger Vanilla Wafer	0.63	8.03	0.34	4.33
Salted Peanuts	0.10	0.42	0.06	0.23
Snickers	-320.00*	-323.42*	-68.06*	-68.79*
Total	-156.02	-73.11	-18.18	25.77

Notes: The left panel reports the change in sales for each product of a simulated removal of Snickers. The simulated profit change of each product for the downstream firm is reported in the right panel. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 12: Simulated Effects of Stockouts on Retailer Profit (Random Coefficients)

	Difference In:	Before Rebate		After Rebate	
Product(s) Removed:	Vends	Margin	Profit	Margin	Profit
Snickers	-73.11	0.98	25.77	0.71	8.19
Animal Crackers	-106.88	-0.48	-70.42	-0.42	-70.24
Doritos	-104.03	-0.05	-43.07	-0.05	-46.88
Peanut M&Ms	-113.89	1.08	8.48	0.78	-10.39
Cheetos	-72.34	-0.01	-28.74	-0.02	-32.09
Choc Chip Famous Amos	-50.20	-0.48	-43.84	-0.42	-42.93
Cheetos + Doritos	-153.59	-0.04	-61.81	-0.06	-68.11
Snickers + Peanut M&Ms	-189.36	2.24	39.22	1.61	0.78

Notes: Difference in margin is reported in cents per unit. Difference in vends is reported in units, and difference in profits is reported in dollars. No profit difference is outside the 90% Confidence Interval described in section 3.1.1.

Table 13: Simulated Effects of Stockouts on Manufacturer Revenue (Random Coefficients)

	Before Rebate						
	Hershey's	Kellogg	Kraft	Mars	PepsiCo	Snyder's	Minor
Snickers	3.55	12.82	3.74	-132.67	13.70	2.31	18.06
Animal Crackers	1.24	-46.43	1.87	15.45	7.64	2.31	6.93
Doritos	0.78	4.62	1.71	7.11	-59.26	2.46	5.26
Peanut M&Ms	3.31	8.41	3.35	-135.83	10.75	1.75	15.54
Cheetos	0.89	8.86	3.05	9.28	-65.59	3.68	9.91
Choc Chip Famous Amos	1.24	-36.23	1.31	15.44	8.22	1.40	5.61
Cheetos + Doritos	1.50	11.32	4.41	14.04	-109.23	5.64	13.40
Snickers + Peanut M&Ms	7.03	23.23	7.27	-283.18	25.72	4.45	37.46
	After Rebate						
Snickers	2.91	11.80	3.45	-111.44	12.33	2.11	17.91
Animal Crackers	1.02	-42.72	1.72	12.98	6.88	2.10	6.84
Doritos	0.64	4.25	1.58	5.98	-53.33	2.24	5.14
Peanut M&Ms	2.72	7.74	3.10	-114.10	9.67	1.60	15.42
Cheetos	0.73	8.15	2.80	7.80	-59.03	3.35	9.62
Choc Chip Famous Amos	1.02	-33.33	1.21	12.97	7.40	1.28	5.54
Cheetos + Doritos	1.23	10.42	4.06	11.79	-98.30	5.14	13.05
Snickers + Peanut M&Ms	5.78	21.37	6.72	-237.86	23.16	4.05	37.20

Notes: Revenues to manufacturer are calculated as the wholesale cost paid by Mark Vend to the manufacturer, not including any potential rebate payments. Revenue impacts for the manufacturer of the focal product of any given experiment are shown in bold typeface (e.g., Mars Inc. manufactures Snickers). Minor manufacturers include all 17 manufacturers listed in table 15.

Table 14: Parameter Estimates and Holdout Analyses

	Base Model	Hold Out Experimental Data from:				
		All Experiments	Candy	Salty Snack	Animal Crackers	Choc Chip F. Amos
<i>Random Coefficients</i>						
σ_{fat}	2.10 (0.091)	2.42	2.39	2.06	2.17	2.14
σ_{salt}	3.49 (0.179)	3.93	4.65	3.21	3.47	3.52
σ_{sugar}	5.25 (0.281)	2.91	2.94	5.44	5.08	5.11
$LL * 10^6$	-5.206	-4.259	-4.806	-4.861	-5.101	-5.111
<i>Nested Logit</i>						
λ_{candy}	0.57 (0.013)	0.66	0.65	0.57	0.58	0.58
λ_{cookie}	0.72 (0.021)	0.74	0.72	0.72	0.73	0.72
λ_{energy}	0.86 (0.015)	0.81	0.84	0.84	0.85	0.86
$\lambda_{salty snack}$	0.62 (0.020)	0.57	0.62	0.60	0.62	0.61
$LL * 10^6$	-5.206	-4.259	-4.806	-4.861	-5.101	-5.111
<i>Number of Choice Sets</i>						
All (417)	1734	1096	1485	1586	1674	1653
Candy (115)	427	262	265	425	427	427
Cookie (53)	184	122	184	184	152	154
Energy (62)	166	166	166	166	166	166
Salty Snack (187)	794	578	792	582	794	793

Notes: Base model is estimated off the full dataset, including all experimental periods. Candy excludes data from the Snickers, Peanut M&M's, and joint Snickers/Peanut M&M Experiments. Salty Snack excludes data from the Cheetos, Doritos, and joint Doritos/Cheetos Experiments. The third panel reports the number of products in each category in parentheses after the category name. The number of unique choice sets is reported in the row labeled "All." The number of choice sets by category is calculated for the products within the category. Category-level choice sets do not sum to the total number of choice sets because variation in choice sets across categories contributes to the total number of choice sets. Standard errors for the base model estimates are reported in parentheses.

Appendix: Manufacturer Statistics and Matching Detail

Table 15 provides summary statistics by manufacturer. The six major snack manufacturers are listed separately, followed by a column for all minor manufacturers.⁶³ In the first two panels, we report statistics for each manufacturer’s full portfolio of products: the inside-good market share, average daily sales per machine, average number of products per machine, and product counts by category. PepsiCo produces the largest number of products (30), with the largest inside market share (over 33%).⁶⁴ The smallest major manufacturer is Kraft, producing 8 products with an inside market share of about 3%. Minor manufacturers account for just over 0.9% of all vends each (15.9% combined), spread across 40 products. The average number of products per machine ranges from 1.6 (out of 8) for Kraft to nearly 11 (out of 30) for PepsiCo. The average daily sales per machine is higher for the major manufacturers than for the combined minor manufacturers, at 1.2 on average. Major manufacturers segment across different product categories, with Mars and Hershey’s focusing on chocolate and candy, PepsiCo and Snyder’s focusing on salty snacks, and Kellogg and Kraft in the cookie and energy categories.

Maximum, minimum and median retail and wholesale prices are listed in the second panel of table 15. The largest spread in prices exists among the combined minor manufacturers. In the last panel, we report retail and wholesale prices and the average inside good share for the “best,” “median,” and “worst” products, defined as the products with the highest/median/lowest average daily sales rate in our three-year sample period. The range of inside good shares is very similar for major and minor manufacturers. Neither retail nor wholesale prices are correlated with sales performance.

Table 16 reports the products used in the matching procedure described in section 3.

⁶³We classify 17 manufacturers as ‘minor’ manufacturers based on the availability and sales of their products at the six experimental sites. These are: Barton’s Confectioners, Biscomerica, Brother’s Kane, California Chips, ConAgra, Farley’s & Sathers Candy Company, Frontera Foods, General Mills, Genisoy, Inventure Group, Just Born Inc., Kar’s Nuts, Nestle, Procter & Gamble, Sherwood Brands, Snak King, and United Natural Foods.

⁶⁴All Frito-Lay and Quaker brands are owned by PepsiCo.

Table 15: Manufacturer Ownership

	Manufacturer						
	Hershey's	Kellogg	Kraft	Mars Inc.	PepsiCo	Snyder's	Minor
<i>Data on Manufacturer Portfolios</i>							
Avg. Inside Good Share	3.36	12.74	4.76	21.11	33.20	8.97	0.93
Avg. Daily Sales per Machine	0.43	1.66	0.62	2.74	4.32	1.16	0.23
Avg. Products per Machine	1.56	5.22	2.77	5.72	10.48	2.80	1.24
# of Products	8	10	14	18	30	10	40
# Chocolate Products	4	0	0	10	0	0	6
# Candy Products	4	0	3	6	0	0	5
# Cookie Products	0	4	7	0	4	0	3
# Energy Products	0	5	4	2	3	0	8
# Salty Snack Products	0	1	0	0	23	10	18
<i>Price Data for Individual Products</i>							
Price (Maximum)	0.85	1.00	0.85	0.85	0.95	1.00	2.00
Price (Median)	0.75	0.75	0.76	0.75	0.74	0.63	0.78
Price (Minimum)	0.75	0.50	0.60	0.60	0.50	0.45	0.50
Wholesale Price (Maximum)	0.57	0.49	0.40	0.67	0.49	0.77	0.77
Wholesale Price (Median)	0.46	0.28	0.28	0.53	0.33	0.23	0.33
Wholesale Price (Minimum)	0.38	0.15	0.16	0.38	0.15	0.14	0.02
<i>Marketshare and Price Data for Individual Products, Based on Performance</i>							
Avg. Inside Share (Best)	5.39	5.33	5.00	6.63	5.85	6.55	5.34
Avg. Inside Share (Median)	1.76	2.48	1.64	1.98	2.93	3.40	1.76
Avg. Inside Share (Worst)	0.02	0.02	0.01	0.09	0.04	1.72	0.03
Price (Best)	0.85	0.85	0.75	0.75	0.50	0.50	0.75
Price (Median)	0.75	0.85	0.85	0.65	0.75	0.75	0.75
Price (Worst)	0.75	0.75	0.85	0.60	0.75	0.60	0.75
Wholesale Price (Best)	0.49	0.39	0.27	0.54	0.23	0.21	0.44
Wholesale Price (Median)	0.49	0.24	0.32	0.54	0.34	0.32	0.02
Wholesale Price (Worst)	0.38	0.28	0.26	0.38	0.28	0.22	0.47

Notes: All calculations performed using the full dataset from Mark Vend Company at six experimental sites. There are 17 minor manufacturers: Barton's Confectioners, Biscomerica, Brother's Kane, California Chips, ConAgra, Farley's & Sathers Candy Company, Frontera Foods, General Mills, Genisoy, Inventure Group, Just Born Inc., Kar's Nuts, Nestle, Procter & Gamble, Sherwood Brands, Snak King, and United Natural Foods. The second panel, "Price data for individual products," ranks products by the reported variable. The third panel ranks products based on their average inside market share and reports statistics for all variables based on this ranking.

Table 16: List of Products Used for Matching at Each Site

<i>Candy and Choc Chip Famous Amos Experiments</i>	
93	Ass. Pretzels, Cheetos, Doritos, Ass. Potato Chips
5055	Cheetos, Doritos
5655	Doritos, Ass. Potato Chips
6056	Ass. Pretzels, Cheetos, Doritos, Ass. Salty Snack
6263	Ass. Pretzels, Cheetos
7277	Ass. Pretzels, Cheetos, Doritos, Ass. Potato Chips, Ass. Salty Snack
<i>Salty Snack Experiments</i>	
93	Ass. Candy, Ass. Chocolate, Raisinets
5055	Raisinets, Twix
5655	Ass. Candy, Ass. Chocolate, Twix
6056	Ass. Candy, Skittles, Twix
6263	Skittles, Twix
7277	Ass. Candy, Ass. Chocolate, Raisinets
<i>Zoo Animal Cracker Experiment</i>	
93	Ass. Pretzels, Ass. Potato Chips, Ass. Salty Snack
5055	Ass. Pretzels, Ass. Potato Chips
5655	Ass. Pretzels, Ass. Potato Chips
6056	Ass. Pretzels, Ass. Potato Chips, Ass. Salty Snack
6263	Ass. Pretzels, Ass. Potato Chips, Ass. Salty Snack
7277	Ass. Pretzels, Ass. Potato Chips

Notes: Candy experiments are Snickers, Peanut M&Ms, and the joint removal of both. Salty Snack experiments are Doritos, Cheetos, and the joint removal of both.