Efficiency and Foreclosure Effects of Vertical Rebates: Empirical Evidence *

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

June 7, 2017

Abstract

Vertical rebates are prominently used across a wide range of industries. These contracts may induce greater retail effort, but may also prompt retailers to drop competing products. We study these offsetting efficiency and foreclosure effects empirically, using data from one retailer. Using a field experiment, we show how the rebate allocates the cost of effort between manufacturer and retailer. We estimate models of consumer choice and retailer behavior to quantify the rebate's effect on assortment and retailer effort. We find that the rebate increases industry profitability and consumer utility, but fails to maximize social surplus and leads to upstream foreclosure.

*We thank Mark Stein, Bill Hannon, and the drivers at Mark Vend Company for implementing the experiment used in this paper, providing data, and generally educating us about the vending industry. We thank Susanto Basu, Michaela Draganska, Scott Hemphill, Wills Hickman, J. F. Houde, Sylvia Hristakeva, Robin Lee, Joe Quinn, Michael Riordan, Greg Shaffer, Mike Whinston, and seminar participants at Bates White, Boston College, Boston University, Harvard Business School, Harvard University (Economics Department), MIT, Pennsylvania State University, Rochester University, SUNY Stony Brook, Texas A & M University, Toulouse School of Economics, Tufts University, Universitat Autonoma de Barcelona, Universitat Pompeu Fabra, University of Toronto, University of Virginia, Washington University in St. Louis, Yale University, and participants at the Quantitative Marketing and Economics conference and the Workshop on Conditional Pricing Practices (held jointly by the Department of Justice and the Federal Trade Commission), for helpful comments. Any remaining errors are our own.

[†]Department of Economics, NYU Stern, 44 W 4th St, New York City, NY 10012. email: cconlon@stern.nyu.edu

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

1 Introduction

Vertical arrangements between manufacturers and retailers have important implications for how markets function. These arrangements may align retailers' incentives with those of manufacturers, and induce retailers to provide demand-enhancing effort. However, they may also reduce competition, exclude competitors, and limit product choice for consumers. Many types of vertical arrangements can induce these offsetting efficiency and foreclosure effects, including resale price maintenance, exclusive dealing, vertical bundling, and rebates, among other contractual forms. Accordingly, these arrangements are a primary focus of antitrust authorities in many countries. Vertical rebates in particular are prominently used across a wide range of industries, including pharmaceuticals, hospital services, microprocessors, snack foods, and heavy industry, and have been the focus of several recent Supreme Court cases and antitrust settlements.[1]

Although vertical rebate contracts are important in the economy and have the potential to induce both pro- and anti-competitive effects, understanding their economic impacts can be challenging. Tension between the potential for efficiency gains on one hand, and exclusion of upstream rivals on the other hand, implies that the contracts must be studied empirically in order to gain insight into the relative importance of the two effects. Unfortunately, the existence and terms of these contracts are usually considered to be proprietary information by their participating firms, frustrating most efforts to study them empirically. An additional challenge for analyzing the effect of vertical contracts is the difficulty in measuring downstream effort, both for the upstream firm and the researcher.

We address these challenges by examining a vertical rebate known as an All-Units Discount (AUD). The specific AUD we study is used by the dominant chocolate candy manufacturer in the United States: Mars, Inc.[2] The AUD implemented by Mars consists of three main features: a retailer-specific per-unit discount, a retailer-specific quantity target or threshold, and a 'facing' requirement that the retailer carry at least six Mars products.

[1] Different forms of vertical rebates include volume-based discounts and 'loyalty contracts.' Volume-based discounts tie payments to a retailer's total purchases from the rebating manufacturer, but do not reference the sales of competing manufacturers. An all-units discount is a particular type of volume-based discount in which the discount is activated once sales exceed a volume threshold. Once activated, the discount applies retroactively to all units sold. We use the term 'loyalty contracts' to refer to payments that are calculated based on a retailer's sales volumes of both the rebating, and competing, manufacturers. Genchev and Mortimer (forthcoming) provides a review of empirical evidence on this class of contracts, including many of the relevant court cases.

[2] With revenues in excess of $50 billion, Mars is the third-largest privately-held company in the United States (after Cargill and Koch Industries).

Mars' AUD stipulates that if a retailer meets the facing requirement and his total purchases exceed the quantity target, Mars pays the retailer an amount that is equal to the per-unit discount multiplied by the retailer's total quantity purchased. We examine the effect of the rebate contract through the lens of a retail vending operator, Mark Vend Company, for whom we are able to collect extremely detailed information on sales, wholesale costs, and contractual terms. The retailer also agreed to run a large-scale field experiment on our behalf, in which we exogenously remove two of Mars' best-selling products and observe subsequent substitution patterns, as well as the profit/revenue impacts for the retailer and all manufacturers. This provides important insight into the effect of the retailer's actions on manufacturer revenues, as well as the potential impact of the AUD on the retailer's decisions. To the best of our knowledge, no previous study has had the benefit of examining a vertical rebate contract using such rich data and exogenous variation.

The insights that we gain from studying Mars' rebate contract allow us to contribute to understanding principal-agent models in which downstream moral hazard plays an important role. Downstream moral hazard arises whenever a downstream firm takes a costly action that is beneficial to the upstream firm but not fully contractible. It is an important feature of many vertically-separated markets, and is thought to drive a variety of vertical arrangements such as franchising and resale price maintenance (RPM).[3] However, empirically measuring the effects of downstream moral hazard is difficult. Downstream effort may be impossible to measure directly, and vertical arrangements are endogenously determined, making it difficult to identify the effects of downstream moral hazard on upstream firms. Our ability to exogenously vary the result of downstream effort (in this case, retail product availability), combined with detailed data on wholesale prices, allows us to directly document the effects of downstream moral hazard on the revenues of upstream firms.

In order to analyze the effect of Mars' AUD contract, we specify a model of consumer choice and a model of retailer behavior, in which the retailer chooses two actions: a set of products to stock, and an effort level. We hold retail prices fixed throughout the analysis, consistent with the data and common practice in this industry.[4] The number of units the

[3]See, among others, Shepard (1993) for an early empirical study of principal-agent problems in the context of gasoline retailing, and Hubbard (1998) for an empirical study of a consumer-facing principal-agent problem.

[4]By holding retail prices fixed, we do not require an equilibrium model of downstream pricing responses to the AUD contract. In practice, we see almost no pricing variation over time or across products within a category (i.e., all candy bars are priced the same as each other, and this price holds throughout the period of analysis). Over a short-run horizon of about three to five years, the retailer typically has exclusive contractual rights to service a location, and some of these contracts may also commit him to a pricing structure during that time.

retailer can stock for each product is constrained by the capacity of his vending machines, and we interpret retailer effort as the frequency with which the retailer restocks his machines. In order to calculate the retailer's optimal effort level, we compute a dynamic restocking model à la Rust (1987), in which the retailer chooses how long to wait between restocking visits.[5] Due to the capacity constraints of a vending machine, the number of unique products the retailer can stock is relatively small. Thus, we compute the dynamic restocking model for several discrete sets of products, and we assume that the retailer chooses to stock the set of products that maximizes his profits. These features of the market (i.e., fixed capacities for a discrete number of unique products) make it well-suited to studying the impacts of the AUD contracts, because the retailer's decisions are discrete and relatively straightforward.[6]

Identification of our consumer choice and supply-side models benefits from two sources of variation. First, we observe a discrete change in the quantity target of Mars' AUD during our sample period. We believe this change was a national change implemented in response to macroeconomic conditions, and not a response to an endogenously determined arrangement with Mark Vend Company. Although re-stocking schedules remain fixed within rebate periods (which are fiscal quarters), we provide evidence that the retailer's re-stocking frequency falls significantly when the quantity target is reduced.

Second, Mark Vend Company implemented a field experiment on our behalf, in which we exogenously vary the stocking decisions. Specifically, the experiment allows us to manipulate the likely outcome of reduced restocking frequency by exogenously removing the best-selling Mars products.[7] The experimental data indicate that in the absence of the rebate contracts, Mars bears almost 90% of the cost of stock-out events. The reason for this is that many consumers substitute to competing brands, which often have higher retail margins. The rebate, which effectively lowers the retailer's wholesale price for Mars products, increases the retailer's share of the cost of stock-out events from around 10% to nearly 50%, and the quantity-target aspect of the rebate provides additional motivation for the retailer to set a high service level.

After estimating the models of consumer choice and retailer behavior, we explore the

[5] Rather than assuming retailer wait times are optimal and using the dynamic model to estimate the cost of re-stocking, we do the reverse: we use an outside estimate of the cost of re-stocking based on wage data from the vending operator, and use the model to compute the optimal wait time until the next restocking visit.

[6] These features also characterize other industries, such as brick-and-mortar retail and live entertainment.

[7] One approach to measuring the impact of effort on profits might be to persuade the retailer to directly manipulate the restocking frequency, but this has some disadvantages. For example, the effects of effort (through decreased stock-out events) are only observed towards the end of each service period, and measuring these effects might prove difficult.

welfare implications of the retailer's optimal effort and assortment decisions. Throughout our counterfactual exercises, our objective is to characterize the set of contracts that lead to foreclosure, and analyze whether or not foreclosure is optimal from a welfare perspective. We do this by checking profit conditions, rather than by calculating a single equilibrium before, and then after, a change to the contracting environment. For example: we show that the retailer prefers to change from carrying two Hershey products with no rebate, to receiving a rebate payment and carrying two Mars products; that Mars prefers to pay the rebate rather than allowing the retailer to carry two Hershey products; and that Hershey lacks a profitable deviation that prevents foreclosure. We focus on the lack of profitable deviations at market outcomes in order to characterize welfare. We then perturb these outcomes (e.g., by changing rebate terms, counterfactual assortments, or ownership), and analyze the impacts for various agents.

Holding retail effort fixed, we examine the incentives created by the AUD for the retailer to change his product assortment. We find evidence that, at the observed wholesale prices and contractual terms, the AUD induces foreclosure of Hershey's products, and we document the role of the quantity target for inducing the foreclosure. We then examine retailer effort, holding assortment fixed. We find evidence that the AUD induces greater retailer effort, and that consumers capture most of the gains from the higher effort level. Following this, we analyze the anticipated changes to producer and consumer surplus that capture both effects (i.e., assortment and effort). We find that social surplus increases, but the market fails to achieve the socially-optimal assortment and effort policies.

In further results, we examine the possibility that Hershey could avoid foreclosure, and find that they cannot. Relatedly, we find that Mars cannot significantly reduce the generosity of its rebate while still foreclosing Hershey. We compare the AUD to uniform wholesale pricing and find that a lower wholesale price would require Mars to transfer a larger share of the rents to the retailer than the AUD does. Finally, we examine the implications of the AUD under various potential upstream mergers. We find, paradoxically, that an AUD can achieve the socially-optimal assortment after a merger between Mars and Hershey. However, we document incentives for the merged firm to reduce the generosity of the AUD post-merger, to the detriment of the retailer.

1.1 Relationship to Literature

There is a long tradition of theoretically analyzing the potential efficiency and foreclosure effects of vertical contracts. The literature that explores the efficiency-enhancing aspects of vertical restraints goes back at least to Telser (1960) and the *Downstream Moral Hazard* problem discussed in Chapter 4 of Tirole (1988).[8] An important theoretical development on the potential foreclosure effects of vertical contracts is the so-called *Chicago Critique* of Bork (1978) and Posner (1976), which makes the point that because the downstream firm must be compensated for any exclusive arrangement, one should only observe exclusion in cases for which it maximizes the profits of the entire industry. Subsequent theoretical literature demonstrates that exclusion may instead maximize industry (or even bilateral) profit, which need not coincide with maximizing efficiency in settings with market power.[9] A separate, but related, theoretical literature has explored the potential anti-competitive effects of vertical arrangements in the context of upfront payments or slotting fees paid by manufacturers to retailers in exchange for limited shelf space (primarily in supermarkets).[10] A broader literature has also examined the conditions under which bilateral contracting might lead to (perhaps partial) exclusion.[11]

Recent theoretical work related to AUDs specifically includes Kolay, Shaffer, and Ordover (2004), which shows that a menu of AUD contracts can more effectively price discriminate than a menu of two-part tariffs when the retailer has private information about demand.[12] More recently, Chao and Tan (2014) show that AUD and quantity-forcing contracts can be used to exclude a capacity-constrained rival, and O'Brien (2013) shows that an AUD may be efficiency enhancing if both upstream and downstream firms face a moral-hazard problem.

We depart from the basic theoretical framework of the *Chicago Critique* of Bork (1978)

[8]In addition, Deneckere, Marvel, and Peck (1996), and Deneckere, Marvel, and Peck (1997) examine markets with uncertain demand and stock-out events, and show that vertical restraints can induce higher stocking levels that are good for both consumers and manufacturers. For situations in which retailers have the ability to set prices, Klein and Murphy (1988) show that without vertical restraints, retailers "will have the incentive to use their promotional efforts to switch marginal customers to relatively known brands...which possess higher retail margins."

[9]For example, Aghion and Bolton (1987) show that long-term contracts that require a liquidated damages payment from the downstream firm to the incumbent can result in exclusion for which industry profits are not maximized; while Bernheim and Whinston (1998) show that the *Chicago Critique* ignores externalities across buyers, and that once externalities are accounted for, it is again possible to generate exclusion that fails to maximize industry profits. Later work by Fumagalli and Motta (2006) links exclusion to the degree of competition in the downstream market. While extremely influential with economists, these arguments have (thus far) been less persuasive with the courts than Bork (1978).

[10]This literature includes Shaffer (1991a) and Shaffer (1991b), which analyze slotting allowances, RPM, and aggregate rebates to see whether or not they help to facilitate collusion at the retail level. Sudhir and Rao (2006) analyze anti-competitive and efficiency arguments for slotting fees in the supermarket industry.

[11]Some key examples include Rasmusen, Ramseyer, and Wiley (1991), Segal and Whinston (2000), and more recently Asker and Bar-Isaac (2014) and Chen and Shaffer (2014).

[12]In addition, Elhauge and Wickelgren (2012) and Elhauge and Wickelgren (2014) explore the potential of loyalty contracts to soften price competition, and Figueroa, Ide, and Montero (2016) examines the role that rebates can play as a barrier to inefficient entry.

and Posner (1976) in some key ways. First, we allow for downstream moral hazard and potential efficiency gains, similar to much of the later theoretical work on vertical arrangements. Second, we study an environment in which the degree of competition across upstream firms may vary across the potential sets of products carried by the retailer, because upstream firms own multiple, differentiated products. Finally, we restrict the retailer to carrying a fixed number of these differentiated products.[13]

The theoretical literature following the *Chicago Critique* focuses on a wide range of settings when considering the potential effects of vertical contracts. Specifically, this literature has studied contracts used by dominant vs. non-dominant firms, contracts that do or do not reference rivals, contracts for which downstream price competition is a major concern for upstream firms (or not), and contracts that apply to single products vs. multiple products. Our setting provides empirical evidence on a vertical rebate used by a dominant firm covering multiple products, for which excessive downstream price competition is not a concern. Although the contract does not explicitly reference rivals, the facing requirement, combined with the typical capacity constraints of most vending machines, effectively limits the presence of competing brands.

One challenge for understanding the effects of vertical arrangements across this wide range of settings is that empirical evidence has primarily been available only through the course of litigation. This has the potential effect that debates about these contracts may be based on a selected sample. An important distinction of our setting is that we study a contract that has not been litigated, and for which we have detailed information on contract terms and exogenous variation in the results of the retailer's effort. Although the welfare effects of vertical rebate contracts in other situations may differ from the impacts we estimate in our setting, we hope that our work provides a road-map for how to model the impacts of these contracts empirically.

Outside of the theoretical literature on vertical rebate contracts, our work also connects to the empirical literature on the impacts of vertical arrangements. One strand of this literature examines issues of downstream moral hazard in the context of vertical integration and the boundaries of the firm, rather than through vertical contracts per se.[14] More recently, another strand of this literature examines exclusive contracts, without necessarily focusing

[13]This contrasts with the "naked exclusion" of Rasmusen, Ramseyer, and Wiley (1991), in which there is a single good.

[14]A few key examples that address downstream (and in some cases upstream) issues of moral hazard include Lafontaine (1992) and Brickley and Dark (1987), which study franchise arrangements, and Baker and Hubbard (2003) and Gil (2007), which study trucking and movies respectively; many other contributions are reviewed in Lafontaine and Slade (2007).

on downstream moral hazard or effort decisions.[15] The most closely-related empirical work is work on vertical bundling in the movie industry, and on vertical integration in the cable television industry. The case of vertical bundling, known as full-line forcing, is studied by Ho, Ho, and Mortimer (2012a) and Ho, Ho, and Mortimer (2012b), which examine the decisions of upstream firms to offer bundles to downstream retailers, the decisions of retailers to accept these 'full-line forces,' and the welfare effects induced by the accepted contracts. The case of vertical integration is studied by Crawford, Lee, Whinston, and Yurukoglu (2015), which examines efficiency and foreclosure effects of vertical integration between regional sports networks and cable distributors. A distinction between our work and Crawford, Lee, Whinston, and Yurukoglu (2015) is that we examine the potential for upstream foreclosure (i.e., manufacturers being denied access to retail distribution), while that study examines the potential for downstream foreclosure (i.e., distributors not having access to inputs).[16]

In practice, AUDs belong to a class of contractual arrangements that the Department of Justice and the Federal Trade Commission refer to as "Conditional Pricing Practices," or CPPs.[17] CPPs are understood to cover any arrangement that allows the terms of sale between a producer and a downstream firm to vary based on whether the downstream firm meets a set of conditions specified by the producer.[18] Genchev and Mortimer (forthcoming) provides a recent survey of empirical evidence on this class of contracts. They find that, "CPPs are more likely to be anticompetitive when dominant firms employ them, when market features force firms to drop competitors' products to comply with the arrangement,

[15]Examples of this literature include Asker (2016), Sass (2005), and Chen (2014), which each examine the efficiency and foreclosure effects of exclusive dealing in the beer industry, and Chipty (2001) and Sinkinson (2014), which study the cable television and mobile phone markets respectively. Lee (2013) focuses on the interaction of exclusive contracts and network effects and competition between downstream firms. Lafontaine and Slade (2008) surveys this literature.

[16]From a methodological perspective, Crawford, Lee, Whinston, and Yurukoglu (2015) differ from us in their use of a bargaining model to describe the equilibrium carriage decisions of cable channels and downstream distributors. These carriage decisions are equivalent to a retailer's choice of product assortment. Both papers model a downstream firm's carriage/stocking decision, given a fixed supply contract, unilaterally as an unobservable (moral hazard) choice. Crawford, Lee, Whinston, and Yurukoglu (2015) employ the bargaining model to help determine supply terms, which we do not model. The biggest difference is that Crawford, Lee, Whinston, and Yurukoglu (2015) examine whether an integrated firm responds to foreclosure incentives in its supply decisions, while we simulate the effects of particular contracts.

[17]The DOJ and the FTC, in June 2014, held a joint workshop exploring the implications of this class of contracts for antitrust policy. Transcripts of that workshop are available at https://www.ftc.gov/news-events/events-calendar/2014/06/conditional-pricing-practices-economic-analysis-legal-policy.

[18]CPPs cover a wide variety of arrangements and are in widespread use throughout many industries. Many of these industries face more complicated decisions than our setting – for example, retailers may be able to restock item by item, choose retail prices, or more perfectly monitor consumer behavior and inventory levels. An outstanding question regarding the use of CPPs is whether or not there are conditions under which one can expect a CPP to be primarily anti-competitive or efficiency inducing.

and when substitute products or alternative distributors are not widely available." While the wide variety of arrangements and the diversity of market structures makes generalization difficult with any observed CPP (including the one we study here), the potential for both anti-competitive and efficiency effects makes it important to build on the empirical body of knowledge about these arrangements. As Genchev and Mortimer (forthcoming) point out, it is especially important to empirically analyze the impacts of CPPs that have not been selected through a process of litigation, to avoid selection bias in the set of contracts examined in the literature.

The rest of the paper proceeds as follows. Section 2 provides the theoretical framework for the model of retail behavior. Section 3 describes the vending industry, data, and the design and results of the field experiment, and section 4 provides the details for the empirical implementation of the model. Section 5 provides results, and section 6 concludes.

2 Theoretical Framework

2.1 Foreclosure and Optimal Assortments: A Motivating Example

We begin by providing a working definition, as well as some examples of the measures of *foreclosure* and *optimal assortment* to be used throughout the rest of our paper. To begin, we focus exclusively on the assortment decision (ignoring effort provision) of the downstream retailer (R) in response to a contract offered by a dominant upstream firm (M). In order to match our empirical application, let us suppose that there are two remaining spaces on the retailer's shelf and the retailer selects from among four potential products (two offered by the dominant firm M, and two offered by the rival H). Let us further assume that both retail prices and wholesale prices are fixed, so that the retailer's sole choice is which products to stock.

We denote the profit of the retailer from stocking the two M products as $\pi^R(M,M)$, from stocking one product from each manufacturer as $\pi^R(H,M)$ and from stocking both products from the rival as $\pi^R(H,H)$. We likewise define the profits of M as $\pi^M(\cdot)$, of the rival H as $\pi^H(\cdot)$, consumer surplus $\pi^C(\cdot)$, and the profits of the industry as $\pi^I(\cdot) = \pi^R(\cdot)+\pi^M(\cdot)+\pi^H(\cdot)$. We define the operator Δ as $\Delta\pi^* = \pi^*(M,M) - \pi^*(H,H)$ for any agent R, M, or H.

Base Case: Two Assortments

In the base case, we assume that the only two possible assortment choices are (M,M) and

(H,H). The dominant firm M offers the retailer R a transfer T in exchange for switching from (H,H) $\rightarrow$ (M,M). In order to make the retailer's decision non-trivial, we assume that $\pi^R(M,M) < \pi^R(H,H)$ (i.e., the retailer earns higher profits when stocking the rival's products).[19] The following conditions (A1)-(A3) ensure that such a transfer is sufficient for M to foreclose its rival H.

(A1) $\Delta\pi^R + T \geq 0$

(A2) $\Delta\pi^M - T \geq 0$

(A3) -$\Delta\pi^H \leq \Delta\pi^M + \Delta\pi^R$

(A1) specifies that the retailer prefers to switch from (H,H) $\rightarrow$ (M,M) after receiving a transfer of size T; (A2), that the dominant firm would be wiling to pay T to induce the retailers to switch from (H,H) $\rightarrow$ (M,M). The third assumption (A3) says that the profits lost by the rival H are smaller than those gained by M and R combined. Thus, (A3) guarantees that even if H offered its own transfer equal to its entire lost profits $\Delta\pi^H(H,H)$, it could not prevent foreclosure.[20]

The ability to obtain foreclosure as an equilibrium outcome is guaranteed by (A3), which may also be restated as $\Delta\pi^I \equiv \Delta\pi^R + \Delta\pi^H + \Delta\pi^M \geq 0$. H is willing to give up all of her profits in order to avoid foreclosure. Thus, when foreclosure is observed, it must be the case that H's losses are smaller than the gains of R and M combined. From the perspective of industry profits, $\Delta\pi^I > 0$, we call this type of foreclosure 'industry optimal.'[21]

Adding a Third Assortment

Now we introduce a new assortment (H,M) which yields intermediate profits for all players:

$$\begin{aligned} \pi^R(H,H) > \quad & \pi^R(H,M) \quad > \pi^R(M,M) \\ \pi^H(H,H) > \quad & \pi^H(H,M) \quad > \pi^H(M,M) \\ \pi^M(H,H) < \quad & \pi^M(H,M) \quad < \pi^M(M,M) \end{aligned} \tag{1}$$

[19]Under an AUD, the transfer would be conditional on meeting a quantity threshold or a facing requirement that is only satisfied under an (M,M) assortment.

[20]If H is fully excluded from the retailer shelf then $\pi^H(M,M) = 0$ and $\Delta\pi^H = \pi^H(H,H)$.

[21]The effect of the change in assortment on consumer surplus $\Delta\pi^C > 0$ or overall social surplus, $\Delta\pi^C + \Delta\pi^I$ may differ from its effect for the industry.

For this case, we ignore the possibility of (M,M), and introduce a new operator $\Delta_H\pi^* = \pi^*(H,M) - \pi^*(H,H)$, with the same set of assumptions:

(B1) $\Delta_H\pi^R + T_h \geq 0$

(B2) $\Delta_H\pi^M - T_h \geq 0$

(B3) $-\Delta_H\pi^H \leq \Delta_H\pi^M + \Delta_H\pi^R$

As above, it is now possible to design a transfer T_h by which M *partially forecloses* the rival H. Again, under (B3) the profits lost by H are less than those gained by the combination of M and R. The resulting (partial) foreclosure is considered feasible in the sense that $\Delta_H\pi^I \equiv \Delta_H\pi^H + \Delta_H\pi^M + \Delta_H\pi^R \geq 0$.

Equilibrium Assortment

If we temporarily ignore the possibility of (H,H), we can consider the effect that the dominant firm's choice of transfer has for obtaining full vs. partial foreclosure and analyze the equilibrium assortment that is obtained when the dominant firm chooses transfers. For this, we introduce a third operator $\Delta_M\pi^* = \pi^*(M,M) - \pi^*(H,M)$ under slightly different assumptions:

(C1) $\Delta_M\pi^R + T_m \geq 0$

(C2) $\Delta_M\pi^M - T_m \geq 0$

(C3) $-\Delta_M\pi^H \leq \Delta_M\pi^M + \Delta_M\pi^R$

(C4) $-\Delta_M\pi^H > \Delta_M\pi^M + \Delta_M\pi^R \geq 0$

(C1) and (C2) are the same as before, but (C3) and (C4) are designed to be mutually exclusive. Either the increase in bilateral surplus among M and R is greater than the losses to H (under (C3)), or it is not (under (C4)). We propose two related results:

Theorem 1. *Under (A1)-(A3), (B1)-(B3) and (C1)-(C3), then there exists a transfer* $T \geq 0$ *such that* (M,M) *is an equilibrium assortment that maximizes industry profits:* $\pi^I(M,M) > \pi^I(H,H)$ *and* $\pi^I(M,M) > \pi^I(H,M)$.

Theorem 2. *Under (A1)-(A3), (B1)-(B3), (C1)-(C2) and (C4), if* $\Delta_M\pi^M + \Delta_M\pi^R \geq 0$, *there exists a transfer* $T \geq 0$ *such that* (M,M) *is an equilibrium assortment even though* $\pi^I(H,M) > \pi^I(M,M) > \pi^I(H,H)$.

Proofs in Appendix A.1.

The main takeaway is that M can set the vector of transfer payments T, T_h, and T_m in order to obtain full (M,M) or partial (H,M) foreclosure. We show that under (A1)-(A3), full foreclosure is feasible.[22] However, if (B1)-(B3) and (C1), (C2), and (C4) also hold, full foreclosure does not lead to the assortment that maximizes overall industry surplus. In this case, partial foreclosure maximizes industry surplus, but full foreclosure leads to higher bilateral surplus among the retailer and dominant firm. As long as the dominant firm chooses the vector of transfers T, T_h and T_m, full foreclosure will be the equilibrium outcome.

The intuition behind this result relates to that of the *Chicago Critique* of Bork (1978) and Posner (1976), which we interpret as asking "When foreclosure is obtained in equilibrium, must the assortment necessarily be optimal?" Our answer is related to the work by Whinston (1990) on tying. When the dominant firm is able to condition the transfer payment on the (M,M) outcome, he can commit to tying the products together, and thus the equilibrium assortment need not maximize the surplus of the entire industry.

2.2 All Units Discount Rebates

In an All Units Discount (AUD) rebate, the transfer T to the retailer is calculated on the basis of all units sold to the retailer, conditional on obtaining a level of sales at or above a required threshold.

Assuming that the dominant manufacturer offers the same wholesale price across all goods w_m and has a constant marginal cost for all goods c_m, one can re-write a quantity-based AUD contract in terms of the profit of the dominant firm, π^M. Denoting the per unit discount payment as d, we define the transfer from M to R as:

$$d \cdot q_m = \underbrace{\left(\frac{d}{w_m - c_m}\right)}_{\lambda} \cdot \pi^M$$

Thus, we denote the payoffs governing the retailer's choice of product assortment as a func-

[22]Furthermore, if the retailer cannot consider an (H,M) assortment, full foreclosure increases overall industry surplus.

tion of M's profits, rather than quantity purchased. Specifically,

$$\begin{cases} \pi^R(a) + d \cdot q_m(a) & \text{if } q_m(a) \geq \overline{q_m} \\ \pi^R(a) & \text{if } q_m(a) < \overline{q_m} \end{cases} = \begin{cases} \pi^R(a) + \lambda \cdot \pi^M(a) & \text{if } \pi^M(a) \geq \overline{\pi^M} \\ \pi^R(a) & \text{if } \pi^M(a) < \overline{\pi^M}. \end{cases}$$

This allows us to define an AUD contract as a tuple $(\lambda, \overline{\pi^M})$, such that, conditional on the dominant firm M receiving a minimum level of profit $\overline{\pi^M}$, the dominant firm transfers a fraction λ of this profit to the retailer.[23] In the notation of the previous section, the transfer is set at $T \equiv \lambda \cdot \pi^M(a)$.

The use of an AUD for setting the transfer has some immediate advantages. If conditions such as those in equation (1) hold with strict inequality, M can tailor the threshold to foreclose the rival by setting $\pi^M(H,M) < \overline{\pi^M} \leq \pi^M(M,M)$. Furthermore, for $\lambda \in [0,1]$, M can transfer between none and all of his profit to R, which means he has access to the full set of transfers that would satisfy conditions such as (A2).[24] In other words, M can write a *de facto* foreclosure contract with an (effectively) unrestricted transfer T. Finally, one can verify the conditions such as (A1)-(A3) at different levels of the AUD terms, $(\lambda, \overline{\pi^M})$, to ask whether (1) the AUD contract can be used to foreclose a rival, (2) the rival could give up her surplus to avoid foreclosure, and (3) the resulting assortment maximizes industry profits.

2.3 Efficiency and Retailer Choice of Effort

A defense of AUD contracts is that they have the potential to be efficiency enhancing if the retailer is encouraged to exert costly effort required to sell the good.[25] This effort can take any number of forms, so long as the effort is costly for the retailer to provide and increases the profits of the dominant firm. When R and M cannot directly contract on the retailer's choice of effort this is known as a *downstream moral hazard* problem (see Tirole (1988) Chapter 4).[26]

In our empirical application, we treat effort as a single scalar variable e which measures the frequency with which the retailer restocks the vending machine. We assume that the

[23] One may alternatively define the threshold in terms of M's remaining profit after the transfer, $(1 - \lambda) \cdot \pi^M(a) \geq (1 - \lambda)\overline{\pi^M}$.

[24] This need not be true for a simpler non-linear pricing schedule. See a comparison to alternative contracts in Appendix A.2.

[25] This defense was employed by Intel in its recent antitrust cases, for example.

[26] Perhaps the best known example is the double marginalization problem. A lower retail price reduces the profits of the retailer, but increases the profits of the wholesale firm (under uniform wholesale pricing and constant marginal cost).

cost of providing effort $c(e)$ is increasing in e. If we hold assortment fixed, the retailer's payoffs under the AUD, as a function of effort, are:

$$\begin{cases} \pi^R(e) - c(e) + \lambda \cdot \pi^M(e) & \text{if } \pi^M(e) \geq \overline{\pi^M} \\ \pi^R(e) - c(e) & \text{if } \pi^M(e) < \overline{\pi^M} \end{cases} \tag{2}$$

The upstream firm can induce greater retailer effort via both features of the contract: (1) a larger per unit discount increases λ so that R gives greater consideration to the profits of M; (2) a larger choice of $\overline{\pi^M}$ leads to greater retailer effort because $\pi^M(e)$ is increasing in effort. In our empirical example, we quantify both of these channels.

We provide a detailed solution to the effort problem in Appendix A.3. To summarize, when effort is non-contractible, R chooses one of three solutions to equation (2): either the interior solution to the effort problem with the rebate (the first line), which we denote e^R, the interior solution to the effort problem absent the rebate (the second line), which we denote e^{NR}, or the solution that makes the constraint bind, $\overline{e} : \pi^M(\overline{e}) = \overline{\pi^M}$. Thus, for $\overline{e} \geq e^R$, M **can set the effort level of the retailer via the threshold** $\overline{\pi^M}$, subject to satisfying the retailer's IR constraint. The set of effort levels that the threshold can target potentially includes the vertically-integrated, and the socially-optimal effort levels. Later, we characterize the critical values of $\overline{\pi}^M$ in our empirical exercise.

An important consideration is whether the potential efficiency gains from increased retailer effort can offset the potential surplus lost from foreclosure. In order to analyze this question, we focus primarily on effort levels that maximize efficiency gains. One can examine the effort choice that is optimal for the bilateral/vertically-integrated firm $M + R$, which we denote e^{VI}, or for the industry (i.e., including profits of the rival), which we denote e^{IND}, or the effort level that maximizes social surplus, denoted e^{SOC}.

We enumerate these possibilities below:

$$\begin{aligned} e^{NR} &= \arg\max_e \pi^R(e) - c(e) \\ e^R &= \arg\max_e \pi^R(e) - c(e) + \lambda \cdot \pi^M(e) \\ e^{VI} &= \arg\max_e \pi^R(e) - c(e) + \pi^M(e) \\ e^{IND} &= \arg\max_e \pi^R(e) - c(e) + \pi^M(e) + \pi^H(e) \\ e^{SOC} &= \arg\max_e \pi^R(e) - c(e) + \pi^M(e) + \pi^H(e) + \pi^C(e) \end{aligned} \tag{3}$$

It may be in the interest of the dominant firm to set a threshold in excess of e^{VI}, because

$\pi^M(e)$ is increasing everywhere. This can be accomplished by choosing a threshold $\overline{\pi^M} > \pi^M(e^{VI})$. For $e < e^{VI}$ the bilateral surplus is increasing in effort, and for $e > e^{VI}$ the bilateral surplus is decreasing in effort; however, at all levels of e, effort (weakly) functions as a transfer from R to M. Thus, in equilibrium, it may be possible to design a transfer that results in socially inefficient excess effort.

3 The Vending Industry and Experimental Data

3.1 Data Description and Product Assortment

We observe data on the quantity and price of all products sold by one retailer, Mark Vend Company. Mark Vend is located in a northern suburb of Chicago, and services 728 snack machines throughout the greater Chicago metropolitan area.[27] Data are recorded internally at each of Mark Vend's machines, and include total vends and revenues since the last service visit to the machine. Any given snack machine can carry roughly 31 standard products at one time. These include salty snacks, cookies, and other products, in addition to 6-8 confection products.[28] We observe retail and wholesale prices for each product at each service visit during a 38-month panel for all snack machines in Mark Vend's enterprise. The dataset covers the period from January, 2006 through February, 2009. There is relatively little price variation over time for any given machine, and almost no price variation within a product category (e.g., confections) for a machine.

A focus in our empirical exercise is the set of products the retailer stocks in the last two slots in the confections category. Mark Vend chooses between stocking two additional Mars products (Milkyway and 3 Musketeers) or two Hershey Products (Reese's Peanut Butter Cups and Payday), or one product from each manufacturer. In table 1 we report the national sales ranks, availability, and shares in the vending industry for the top-ranked products nationally, as well as the availability and shares for the same products at Mark Vend's machines. There are some patterns that emerge. The first is that Mark Vend stocks some of the most popular products sold by Mars (Snickers, Peanut M&Ms, Twix, Plain M&M's, and Skittles) in most of his machines. However, Mark Vend only stocks Hershey's best-selling product (Reese's Peanut Butter Cups) in 27% of machine-weeks, even though nationally Reese's Peanut Butter Cups is the fourth most popular product. Overall Mark Vend tends to sell more Mars products (around 73% of all confections sales) than the national average

[27]Mark Vend services an additional 800+ machines that vend beverages, frozen food, or coffee machines.
[28]Most machines have another 4-5 slots for smaller items, such as gum and mints.

(around 52% of all confections sales). The non-Mars product most frequently stocked by Mark Vend is Nestle's Raisinets (at 47% of machine-weeks), which does not rank in the top 45 products nationally in confections sales.

There are two possible explanations for Mark Vend's departures from the national best-sellers. One is that Mark Vend has better information on the tastes of its specific consumers, and its product mix is geared towards those tastes. The alternative explanation is that the rebate induces Mark Vend to substitute from Nestle/Hershey brands to Mars brands when making stocking decisions, and that when Mark Vend does stock products from competing manufacturers (e.g., Nestle Raisinets), he chooses products that do not steal business from key Mars products.

3.2 Mars' AUD with Mark Vend

Mars' AUD rebate program is the most commonly-used vertical arrangement in the vending industry.[29] Under the program, Mars refunds a portion of a vending operator's wholesale cost at the end of a fiscal quarter if the vending operator meets a quarterly sales goal. The sales goal for an operator is typically set on the basis of its combined sales of Mars' products, rather than for individual Mars products. Mars' rebate contract also stipulates a minimum number of product 'facings' that must be present in an operator's machines, although in practice, this provision is difficult to enforce because Mars cannot observe the assortments in individual vending machines. The amount of the rebate and the precise threshold of the sales goal are specific to an individual vending operator, and these terms are closely guarded by participants in the industry.

We include some promotional materials from Mars' rebate program in figure 1.[30] The program employs the slogan *The Only Candy You Need to Stock in Your Machine!*, and specifies a facing requirement of six products and a quarterly sales target. The second page of the document shown in figure 1 refers to discontinuing a growth requirement, which we

[29] For confections products, Mars is the dominant manufacturer in vending, and is the only manufacturer to offer a true AUD contract. The AUD is the only program offered to vendors by Mars. Hershey and Nestle offer wholesale 'discounts,' but these have a quantity threshold of zero (i.e., their wholesale pricing is equivalent to linear pricing). The salty snack category is dominated by Frito-Lay (a division of PepsiCo) which does not offer a rebate contract. We do not examine beverage sales, because many beverage machines at the locations we observe are serviced directly by Coke or Pepsi.

[30] A full slide deck, titled '2010 Vend Program,' and dated December 21, 2009, is available at http://vistar.com/KansasCity/Documents/Mars%202010%20Operatopr%20rebate%20program.pdf. (Last accessed on April 19, 2015; available from the authors upon request.) These promotional materials represent the same type of rebate in which Mark Vend participated, but may differ from the terms available to Mark Vend during the period we study.

believe to be 5% (i.e., a target of 105% of year-over-year sales). On another page, not shown in figure 1, the document describes the sales target for a "Gold" rebate level as 90% of year-over-year quarterly sales. The rebate does not explicitly condition on market share or the sales of competitors. However, most vending machines typically carry between six and eight candy bar varieties, so the facing requirement may effectively limit shelf space for competing brands.[31]

We observe, but cannot report, the amount of the rebate received by Mark Vend Company. However, we can construct quarterly sales of Mars products at Mark Vend, and compare the year-over-year sales across Mark Vend's entire enterprise for all but the first four quarters of our data. We present those calculations in table 2. We see that from the first quarter of 2007 through the first quarter of 2008, Mark Vend generally hits a threshold of 105% of year-over-year sales. (The exception is the third quarter of 2007, when he sells 100% of year-over-year sales.)

In the wake of the 2008 macroeconomic downturn, Mars modified its rebate program and reduced the threshold. We can see clearly in table 2 that Mark Vend's sales of Mars products appear to respond to the lower threshold, and indeed track a 90% threshold quite closely.[32] This response comes primarily through a lower share of Mars products (declining from 20-21% in the third quarter of 2007 down to 17.6% in the first quarter of 2009). At the same time, we see that Mark Vend's enterprise-level sales were not hit particularly badly by the macroeconomic downturn, as (normalized) total vends across all products remained largely flat between 2007 and 2009.

Under the assumption that the reduction in Mars' rebate threshold is an exogenous event (rather than a direct response to behavior by Mark Vend), we can examine its impact on Mark Vend's assortment and effort decisions. To examine the impact on assortment, we count the average number of product facings per machine dedicated to each manufacturer's products. In table 3, we see that when Mars reduced the threshold, around the third quarter of 2008, Mark Vend reduced the number of Mars product facings in an average vending machine

[31] While there is some ability for a vending operator to adjust the overall number of candy bars in a machine, it is often difficult to do without upgrading capital equipment, because candy bars and salty snacks do not use the same size 'slots.'

[32] Our data reflect retail sales in vending machines, while the sales targets are derived from wholesale cases ordered. In later analyses, we implicitly assume that retail sales track wholesale orders perfectly. Some products may spoil or melt, or be damaged in delivery or stolen. Likewise, the retailer can place his wholesale orders in order to meet the threshold, while holding extra inventory in his warehouse (or even disposing of products). This implies there is a small margin of error between our threshold calculation and the calculation Mars uses to establish whether the conditions of the rebate have been met. In correspondence with Mark Vend, they assure us that these effects are small and do not change over time.

from around 6.6 to around 5.3. Over the same time period, the number of Hershey facings increased from around 1 facing per machine to around 2 facings per machine. The right-hand-side panel of the table shows that the major switch was to swap Mars' Three Musketeers (stocked in around half of machines at the beginning of the sample) for Hershey's Reese's Peanut Butter Cups and Payday (stocked in 62% and 23% of machines respectively at the end of the sample period). Although it is difficult to attribute causality, it is worth pointing out that prior to the reduction in the threshold, both Reese's Peanut Butter Cups and Payday are effectively foreclosed, as they are stocked in very few of Mark Vend's machines.

We can also measure how Mark Vend adjusts his effort when the sales threshold changes. In table 4, we report regression results at the machine-visit level for two effort variables: the number of vends between visits and the elapsed number of days between visits. We include machine and week-of-year fixed effects. Thus, the regressions examine variation in these effort variables within a particular vending machine over time, while trying to control for overall seasonality in how often machines are serviced (if, for example, employees in office buildings take more vacation in summer). We find that after the threshold is reduced (in the third quarter of 2008), Mark Vend waits an average of 0.85 days longer before servicing machines, and that machines have sold 8.26 more products on average since the last service visit. Together, these imply that Mark Vend is reducing effort, rather than merely responding to a slower rates of sales. While one must be cautious about causally interpreting the retailer's response to changes in the threshold by Mars, it appears that there is both a substantial reduction in his "effort," as measured by service frequency and sales between visits, and a substantial change in assortment, based on the information in table 3.

3.3 Experimental Design

When we run our experiment and estimate our consumer choice model, we focus on a set of 66 vending machines that are located in high-income, professional office environments in Chicago, where consumers may have very different tastes than consumers from other demographic groups.[33]

In addition to sharing the terms of his rebate contact with us, the owner of Mark Vend implemented a field experiment for us in which his drivers exogenously removed either one or two top-selling Mars confection products from the set of 66 'experimental' machines. The

[33] For example, Starburst, a fruit flavored candy sold by Mars, is primarily marketed to younger consumers. It is stocked far more often across Mark Vend's entire enterprise (around 41% of machine-visits) when compared to our sample of experimental machines (16%).

product removals are recorded during each service visit.[34] Implementation of each product removal was fairly straightforward; the driver removed either one or both of the two top-selling Mars products from all machines for a period of roughly 2.5 to 3 weeks. The focal products were Snickers and Peanut M&Ms.[35] The dates of the product removal interventions range from June 2007 to September 2008, with all removals run during the months of May - October. Over all sites and months, we observe 185 unique products. We consolidate products that had very low levels of sales with similar products within a category that are produced by the same manufacturer, until we are left with the 73 'products' that form the basis of the rest of our exercise.[36]

During each 2-3 week product removal period, most machines receive about three service visits. However, the length of service visits varies across machines, with some machines visited more frequently than others. Machines are serviced on different schedules, and as a result, it is convenient to organize observations by machine-week, rather than by visit, when analyzing the results of the experiment. When we do this, we assume that sales are distributed uniformly among the business days in a service interval, and assign those business days to weeks. Different experimental treatments start on different days of the week, and we allow our definition of when weeks start and end to depend on the client site

[34]The machines are located in office buildings, and have substitution patterns that are very stable over time. In addition to the three treatments described here, we also ran five other treatment arms, for salty-snack and cookie products, which are described in Conlon and Mortimer (2010) and Conlon and Mortimer (2013b). The reader may refer to our other papers for more details.

[35]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, Mark Vend wanted to minimize the number of phone calls received in response to the stock-out events. 'Natural,' or non-experimental, stock-outs are extremely rare for our set of machines. This implies that much of the variation in product assortment comes either from product rotations, or our own exogenous product removals. Product rotations primarily affect 'marginal' products, so in the absence of exogenous variation in availability, the substitution patterns between marginal products is often much better identified that substitution patterns between continually-stocked best-selling products. Conlon and Mortimer (2010) provides evidence on the role of the experimental variation for identification of substitution patterns.

[36]For example, we combine Milky Way Midnight with Milky Way, and Ruffles Original with Ruffles Sour Cream and Cheddar. In addition to the data from Mark Vend, we also collect data on product characteristics online and through industry trade sources. For each product, we note its manufacturer, as well as the following set of product characteristics: package size, number of servings, and nutritional information. Nutritional information includes weight, calories, fat calories, sodium, fiber, sugars, protein, carbohydrates, and cholesterol. For consolidated products, we collect data on product characteristics at the disaggregated level. The characteristics of the consolidated product are computed as the weighted average of the characteristics of the component products, using vends to weight. In many cases, the observable characteristics are identical.

and experiment.[37]

Two features of consumer choice are important for determining the welfare implications of the AUD contract. These are, first, the degree to which Mark Vend's consumers prefer the marginal Mars products (Milky Way, Three Musketeers, Plain M&Ms) to the marginal Hershey products (Reese's Peanut Butter Cup, Payday), and second, the degree to which any of these products compete with the dominant Mars products (Peanut M&Ms, Snickers, and Twix). Our experiment mimics the impact of a reduction in retailer effort (i.e., restocking frequency) by simulating the stock-out of the best-selling Mars confections products. This provides direct evidence about which products are close substitutes, and how the costs of stock-outs are distributed throughout the supply chain. It also provides exogenous variation in the choice sets of consumers, which helps to identify the discrete-choice model of consumer choice.

In principle, calculating the effect of product removals is straightforward. In practice, however, there are two challenges in implementing the removals and interpreting the data generated by them. First, there is variation in overall sales at the weekly level, independent of our exogenous removals. Second, although the experimental design is relatively clean, the product mix presented in a machine is not necessarily fixed across machines, or within a machine over long periods of time, and we rely on observational data for the control weeks. To mitigate these issues, we report treatment effects of the product removals after selecting control weeks to address these issues. We provide the details of this procedure in Appendix A.4.

3.4 Results of Product Removals

Our first exogenous product removal eliminated Snickers from all 66 vending machines involved in the experiment; the second removal eliminated Peanut M&Ms, and the third eliminated both products.[38] These products correspond to the top two sellers in the confections category, both at Mark Vend and nationwide.

One of the results of the product removal is that many consumers purchase another product in the vending machine. While many of the alternative brands are owned by Mars, several of them are not. If those other brands have similar (or higher) margins for Mark Vend, substitution may cause the cost of each product removal to be distributed unevenly across the supply chain. Table 5 summarizes the impact of the product removals for Mark

[37] For example, at some site-experiment pairs, we define weeks as Tuesday to Monday, while for others we use Thursday to Wednesday.

[38] As noted in table 1, both Snickers and Peanut M&Ms are owned by Mars.

Vend. In the absence of any rebate payments, we see the following results. Total vends decrease by 217 units and retailer profits decline by \$56.75 when Snickers is removed. When Peanut M&Ms is removed, vends go down by 198 units, but Mark Vend's average margin on all items sold in the machine rises by 0.78 cents, and retailer revenue declines only by \$10.74 (a statistically insignificant decline). Similarly, in the joint product removal, overall vends decline by roughly 283 units, but Mark Vend's average margin rises by 1.67 cents per unit, so that revenue declines by only \$4.54 (again statistically insignificant).[39]

Table 6 examines the impact of the product removals on the upstream firms. Removing Peanut M&Ms decreases Mars' revenue by about \$68.38, compared to Mark Vend's loss of \$10.74; thus roughly 86.4% of the cost of stocking out is born by Mars (reported in the fifth column). In the double removal, because Peanut M&M customers can no longer buy Snickers, and Snickers customers can no longer buy Peanut M&Ms, Mars bears 96.7% of the cost of the stockout. In the Snickers removal, most of the cost appears to be born by the downstream firm; one potential explanation is that among consumers who choose another product, many select another Mars Product (Twix or Peanut M&Ms). We also see the impact of each product removal on the revenues of other manufacturers. Hershey (which owns Reese's Peanut Butter Cups and Hershey's Chocolate Bars) enjoys relatively little substitution in the Snickers removal, in part because Reese's Peanut Butter cups are not available as a substitute. In the double removal, when Peanut Butter Cups are available, Hershey profits rise by nearly \$61.43, capturing about half of Mars' losses. We see substitution to the two Nestle products in the Snickers removal, so that Nestle gains \$19.32 as consumers substitute to Butterfinger and Raisinets; Nestle's gains are a smaller percentage of Mars' losses in the other two removals.

Direct analysis of the product removals can only account for the marginal cost aspect of the rebate (i.e., the price reduction given by λ); one requires a model of restocking in order to account for the threshold aspect, $\overline{\pi^M}$. By more evenly allocating the costs of stocking out, the rebate should better align the incentives of the upstream and downstream firms, and lead the retailer to increase his overall service level. Returning to table 5, the right-hand panel reports the retailer's profit loss from the product removals after accounting for his rebate payments, assuming he qualifies. We see that the rebate reallocates approximately (\$17, \$30, \$50) of the cost of the Snickers, Peanut M&Ms, and joint product removals from the upstream to the downstream firm. The last column of table 6 shows that after accounting

[39] Total losses appear smaller in the double-product removal in part because we sum over a smaller sample size of viable machine-treatment weeks (89) for this experiment, compared to the Peanut M&Ms removal (with 115 machine-treatment weeks).

for the rebate contract, the manufacturer bears about 50% of the cost of the Peanut M&Ms removal, 60% of the cost of the joint removal, and 12% of the cost of the Snickers removal.

4 Estimation

4.1 Consumer Choice

In order to consider the optimal product assortment, we need a parametric model of consumer choice that predicts sales for a variety of different product assortments. We estimate a mixed (random-coefficients) logit model on our sample of 66 machines (including both experimental and non-experimental periods).[40]

We consider a model of utility in which consumer i receives utility from choosing product j in market t of:

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}. \tag{4}$$

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. Each consumer has an outside option $u_{ijt} = \varepsilon_{i0t}$ of "no-purchase", which includes the possibility of not having a snack, bringing a snack from home, or purchasing a snack from somewhere other than a vending machine.

A random-coefficients logit specification allows for correlation in tastes across observed product characteristics.[41] 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:

$$s_{jt}(\delta,\theta,a_t) = \int \frac{e^{\delta_{jt}+\sum_l \sigma_l \nu_{ilt} x_{jl}}}{1+\sum_{k\in a_t} e^{\delta_{kt}+\sum_l \sigma_l \nu_{ilt} x_{kl}}} f(v_{ilt}|\theta). \tag{5}$$

We define a_t as the set of products stocked in market t, and a market as a machine-visit pair (i.e., a_t is the product assortment stocked in a machine between two service visits). We

[40] Results from an alternative nested-logit specification are available from the authors upon request.
[41] See Berry, Levinsohn, and Pakes (1995).

estimate the potential daily market size for each machine, $\hat{M}_t$, as twice the maximum daily sales rate observed at the machine across our panel.

There are virtually no 'natural' stock-outs in the data; thus, changes to product assortment happen for two reasons: (1) Mark Vend changes the assortment when re-stocking, or (2) our field experiment exogenously removes one or two products. While Mark Vend's assortment decisions are chosen endogenously, they are often temporary and due to changes in manufacturer product lines.[42] There is considerable product churn created by non-experimental changes in assortment, which helps to identify substitution between non-focal products. Non-experimental churn creates 262 unique choice sets for confection products; our exogenous product removals increase the number of unique choice sets to 427.[43]

Implicitly, our estimation of the consumer choice model assumes away dynamic effects of stock-outs (i.e., we assume no change in consumer preferences after the temporary removal of a product).[44] Nevertheless, one should view our consumer choice model as capturing substitution patterns that are stable in the short run. Other factors (including manufacturer advertising) may impact substitution patterns in the long run.

We specify $\delta_{jt} = d_j + \xi_t$; that is, we allow for 73 product intercepts as well as market-specific demand shifters. We allow for three random coefficients, corresponding to consumer tastes for salt, sugar, and nut content.[45] We estimate the parameters of the choice probabilities via maximum simulated likelihood (MSL). The log-likelihood is:

$$l_t(\mathbf{y_t}|\delta, \theta, a_t) \quad \propto \quad \sum_j y_{jt} \log s_j(\delta, \theta, a_t). \tag{6}$$

where y_{jt} are sales of product j in market t.

We report the parameter estimates in table 7. We report two levels of aggregation for

[42] Implicitly, we assume that changes to manufacturer product lines are taken with the national market in mind, rather than to induce a behavioral change by Mark Vend.

[43] Further discussion and analyses of choice-set variation in this dataset are contained in Conlon and Mortimer (2010).

[44] Using the same data, Kapor (2008) examines this assumption and finds no evidence that temporary stock-outs affect future demand patterns.

[45] Nut content is a continuous measure of the fraction of product weight that is attributed to nuts. We 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 (2013a), 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 estimated a number of alternative specifications in which we included random coefficients on other continuous variables, such as carbohydrates, fat, or calories. In general, the additional parameters were not significantly different from zero, and they had no appreciable effect on the results of any prediction exercises.

ξ_t. The first allows for 15,256 fixed effects, at the level of a machine-service visit, while the second allows for 2,710 fixed effects, at the level of a machine-choice set (i.e., we combine machine-service visit 'markets' for which the choice set does not change). We report the log-likelihood, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for each specification. We use BIC to select the specification with 2,710 ξ_t fixed effects. Our simulated MLE parameters tend to be very precisely estimated, because we observe 2.96 million sales.

Parametric identification of d_j and σ parameters is straightforward. The d_j parameters would be identified from average sales levels in even a single market after we normalize the utility of the outside good to zero. Across machines and time, we observe 2,710 different product assortments a_t. The σ parameters are identified by the covariance of the changes in the observed sales across product assortments with the characteristics of the products that are added or removed from the choice set. For example, when we exogenously remove Peanut M&Ms during our experiment, we observe whether more consumers appear to switch to products with a similarly high peanut content (such as Planter's Peanuts) or to products with a similar sugar content (such as Plain M&Ms). A common identification challenge in the literature is the identification of an (endogenous) price effect. In our application, price effects are subsumed into d_j because we do not observe any within-product price variation (the entire confections category is priced at 75 cents in our sample).

4.2 Supply

On the supply side, we begin with the retailer's problem, taking the manufacturer's choice of contract terms as given. We model the retailer's optimal choices of assortment, a, and effort, e. Once we characterize the optimal retail choices at the existing contract terms, $(\lambda, \overline{\pi^M})$, we can re-solve the retailer's problem at different values for $(\lambda, \overline{\pi^M})$. In this sense, our exercise does not use the supply-side model to characterize a single equilibrium before, and then after, a change to a competitive environment in order to calculate a welfare effect. Instead, we use information about the retailer's supply-side decisions to calibrate a supply-side model and explore (via simulation) the space of contract terms. This allows us to analyze the values of the contract terms that may lead to foreclosure, and to quantify the welfare effects of different contract terms for various agents in the industry.

The retailer's problem is:

$$\max_{a,e} \begin{cases} \pi^R(a,e) - c(e) + \lambda \cdot \pi^M(a,e) & \text{if } \pi^M(a,e) \geq \overline{\pi^M} \\ \pi^R(a,e) - c(e) & \text{if } \pi^M(a,e) < \overline{\pi^M} \end{cases} \tag{7}$$

where $\pi^R(a,e)$ is the variable profit of the retailer absent any rebate payment, $\pi^M(a,e)$ is the variable profit of the dominant manufacturer M, and $c(e)$ represents the cost of retailer effort.

The retailer's assortment decision involves simple discrete comparisons across a finite number of choices. We explain the set of potential assortments that we analyze in section 4.2.3. For each potential choice of assortment, we calculate the retailer's optimal choice of effort.

4.2.1 Retail Effort Choice: Dynamic Model of Re-stocking

We believe that Mark Vend's effort decision is operationalized as follows. At the beginning of each quarter, MarkVend decides on an (enterprise-wide) policy to restock after e likely consumers have arrived at all of his vending machines.[46] He then translates this policy into a restocking schedule for each individual vending machine (e.g., every Tuesday, every 10 days, every other day, etc.) based on knowledge of a machine-specific arrival rate. Once the schedule for the quarter is set, he breaks up the schedule into individual service routes, and assigns routes to drivers and trucks. In order to reduce the number of consumer arrivals between service visits, MarkVend must hire additional trucks and drivers, which increases his costs. An implication of this setup is that MarkVend commits to a restocking policy for an entire quarter. This means that if sales are below expectations (i.e., if he repeatedly draw from the left-tail of the consumer arrival distribution), MarkVend does not adjust his stocking policy until the next quarter.[47]

In our application, we consider the specific case in which the retailer chooses the restocking frequency. We model the retailer's choice of effort, e, using an approach similar to Rust

[46] Mars' AUD rebate contract is evaluated quarterly on the basis of MarkVend's entire enterprise, which includes 728 snack vending machines.

[47] Within a quarter, it appears as the most machines are on an extremely predictable fixed schedule, and there is no evidence that the schedule is adjusted in either direction towards the end of each quarter. This is consistent with a model of effort in which the frequency of service is set in response to the payoff function, but the schedule is not set dynamically within a quarter as a function of the distance from the threshold. As Mark Vend does not observe sales, except at the time of a service visit, this makes a lot of sense. He doesn't have new information by which to dynamically adjust a service schedule across days.

(1987), but 'in reverse.' Rather than assuming that observed retailer wait times are optimal and using Rust's model to estimate the cost of re-stocking, we use an outside estimate of the cost of re-stocking based on wage data from the vending operator, and use the model to compute the optimal wait time until the next restocking visit. In order to model the choice of effort, we consider a multi-product (**s**,**S**) policy, in which the retailer pays a fixed cost FC and fully restocks (all products) to target inventory **S**. The challenge is to characterize the critical re-stocking inventory level, **s**. In our application, it is more convenient to work with the number of potential consumer arrivals (a scalar, which we denote x), rather than working with the vector **s**.[48] This implies an informational restriction on the retailer: namely, that he observes the number of potential consumers (for example, the number of consumers who pass by a vending machine), but not necessarily the actual inventory levels of each individual product, when he makes his restocking decision. This closely parallels the problem of Mark Vend.[49]

Mark Vend solves the following dynamic stocking problem, where $u(x)$ denotes the cumulative variable retailer profits after x potential consumers have arrived. Profits are not collected by Mark Vend until he restocks. His value function is:

$$V(x) = \max\{u(x) - FC + \beta V(0), \beta E_{x'}[V(x'|x)]\}. \tag{8}$$

The problem posed in equation (8) is similar to the 'Tree Cutting Problem' of Stokey, Lucas, and Prescott (1989), which for concave $u(x)$ and increasing $x' \geq x$, admits a monotone policy such that the firm re-stocks if $x \geq x^*$. Given a guess of the optimal policy, we can compute the post-decision transition-probability-matrix $\tilde{P}$ and the post-decision pay-off $\tilde{u}$, defined as:

$$\tilde{u}(x,x^*) = \begin{cases} 0 & \text{if } x < x^* \\ u(x) - FC & \text{if } x \geq x^*. \end{cases}$$

This allows us to solve the value function at all states in a single step:

$$V(x,x^*) = (I - \beta \tilde{P}(x^*))^{-1} \tilde{u}(x,x^*). \tag{9}$$

[48]In the multi-product setting, **s** is multi-dimensional (and may not define a convex set).

[49]That is, Mark Vend may have information on whether particular days are likely to be busy or not, but does not observe the actual inventory levels of individual products until visiting the machine to restock it. In other retail contexts this assumption might be less realistic and could be relaxed; its role is primarily to reduce the computational burden in solving the re-stocking problem.

This also enables us to evaluate profits under alternative stocking policies x', or policies that arise under counterfactual market structures. For example, in order to understand the incentives of a vertically-integrated firm, $M + R$, we can replace $u(x)$ with $(u^R(x) + u^M(x))$, which incorporates the profits of the dominant upstream manufacturer. Likewise, we can consider the industry-optimal policy by replacing $u(x)$ with $(u^R(x)+u^M(x)+u^H(x)+u^N(x))$.

To find the optimal policy we iterate between (9) and the policy improvement step:

$$x^* = \min x : u(x) - FC + \beta V(0,x^*) \geq \beta P(x'|x)V(x',x^*). \tag{10}$$

The fixed point $(x^*,V(x,x^*))$ maximizes the long-run average profit of the agent $\Gamma(x^*)V(x,x^*)$ where $\Gamma\tilde{P} = \Gamma$ is the ergodic distribution corresponding to the post-decision transition matrix. These long-run profits will become the basis on which we compare contracts and product assortment choices.

4.2.2 Retail Effort Choice: Empirical Implementation

In order to compute the dynamic restocking model, we construct a 'representative vending machine' via the following procedure. We define a 'full machine' as one that contains a set of the 29 most commonly-stocked products, which we report in table 8, and we use actual machine capacities for each product.[50] Beginning with a full machine, we simulate consumer arrivals one at a time and allow consumers to choose products in accordance with the mixed logit choice probabilities $s_{jt}(\delta,\theta,a_t)$ (including an outside option of no-purchase). After each consumer choice, we update the inventories of each product and adjust the set of available products a_t if a product has stocked out. When products stock out, consumers substitute to other products, including the no-purchase option. We continue to simulate consumer arrivals until the vending machine is empty. We average over 100,000 simulated chains to construct the expected profits after x consumers have arrived, and fit a smooth Chebyshev polynomial to the profits of each agent $\hat{u}^R(x),\hat{u}^M(x),\hat{u}^H(x),\hat{u}^C(x)$.[51]

The state variable of our dynamic programming problem, X_t, is the number of potential consumers who have arrived since our 'representative vending machine' was last restocked. The exogenous state transition matrix $P(X_{t+1} - X_t|X_t) \approx P(\Delta X_t)$ is the incremental number of potential consumers who arrive to the representative vending machine each business day.

[50] These capacities are nearly uniform across machines, and are: 15-18 units for each confection product, 11-12 units for each salty snack product, and around 15 units for each cookie/other product.

[51] The fit of the 10th order Chebyshev polynomial is in excess of $R^2 \geq 0.99$. It is generally well behaved except at the very edges of the state space, but these are far from our optimal policies.

We assume that the arrival rate has a discrete distribution.[52] In a separate stage, we use 28 of our 66 experimental machines to form a non-parametric estimate of $P(\Delta x)$. These 28 machines have an average daily sales volume of 15.1 units and a standard deviation of 2.0 units.[53] For each service-visit observation at each of these machines, we use the number of estimated consumer arrivals since the last service visit, and divide this by the number of elapsed business days since the last visit to compute the number of daily consumer arrivals, Δx_t.[54] Effort policies are not particularly sensitive to the specification of the arrival process.[55]

We choose a daily discount factor $\beta = 0.999863$, which corresponds to a 5% annual interest rate. We assume a fixed cost of a restocking visit, $FC = \$10$, which approximates the per-machine restocking cost using the driver's wage and average number of machines serviced per day. As a robustness test, we also consider $FC = \{5,15\}$, which generate qualitatively similar predictions. In theory, one should able to estimate FC directly off the data using the technique of Hotz and Miller (1993). However, our retailer sets a level of service that is too high to rationalize with any optimal stocking behavior, often refilling a day before any products have stocked-out.[56] This is helpful as an experimental control, but makes identifying FC from data impossible.

In order to speed up computation, we normalize our state space when solving the dynamic programming problem. Instead of working with the number of consumers to arrive at

[52] This mimics Rust (1987) who estimates a discrete distribution of weekly incremental mileage.

[53] The machines in this group have higher than average sales volumes, but are not the largest machines. We chose this group for our exercise because we think it is the most important set of machines for determining the retailer's re-stocking decision. For additional detail, please see Appendix A.5.

[54] Note that the data report average daily sales, rather than consumer arrivals (i.e., there are no cameras on the vending machines). As in the consumer choice model, the relationship between observed sales and consumer arrivals depends on availability. If a machine is empty, no sales will occur, regardless of the consumer arrival rate. The consumer choice model adjusts for this by allowing substitution to remaining products (including the outside good) when a machine is not fully stocked. Our estimate of consumer arrivals uses the same adjustment.

[55] Doubling or tripling the rate at which consumers arrive has very little effect on the optimal effort policy, because policies are defined in terms of the cumulative number of consumer arrivals x (rather than days, for example). In robustness tests, we assume that the firm can make decisions consumer-by-consumer, or only every four 'days.' With appropriate scaling of the discount factor β, the optimal policies change by only 2-3 units.

[56] In conversations with the retailer about his service schedule, he provided two explanations of this fact. First, he suspected that he was over-servicing, and reduced service levels after our field experiment. Second, he explained that high service levels are important to obtaining long-term (3-5 year) exclusive service contracts with locations. Our specific experimental locations almost certainly do not reflect a company-wide servicing policy. Specifically, these are high-end office buildings with high service expectations. Public locations, such as museums and hospitals, have much higher levels of demand and higher rates of stock-out events. These public locations affect company-wide servicing policies, but are not good candidates for running a successful field experiment.

the vending machine, we work with the number of consumers who would have likely made a purchase at a hypothetical 'full' vending machine. This saves us from simulating large numbers of consumers who always choose the outside good, independent of product assortment. We thus label our state-space as 'likely' consumer arrivals instead of 'potential' consumer arrivals from this point forward.[57]

By simulating from our consumer choice model in section 4.1, we can compute the payoffs to each agent from any assortment a and any effort level e using equation (9). For the retailer, with effort policy e:

$$\pi^R(a,e) = \Gamma(\tilde{P}(e)) \cdot (I - \beta\tilde{P}(e))^{-1} \cdot \hat{u}^R(x,a), \tag{11}$$

and represents the net present value of the long-run average (infinite horizon) profits of a single representative vending machine under assortment a and restocking policy e.[58] The model express effort as a measure of service frequency, for which the policy function answers the question "After how many likely consumers should Mark Vend re-stock?" Therefore, if MarkVend restocks a vending machine after 240 consumers instead of after 260 consumers, he is restocking more often and exerting more effort. In order to evaluate profits for different agents, we replace $\hat{u}^R(x,a)$ with the profit of the relevant agent and evaluate at the same effort policy (i.e., we replace $\hat{u}^R(x,a)$ with $\hat{u}^M(x,a)$ to evaluate the profits of the dominant retailer Mars).

4.2.3 Retail Assortment Choice: Empirical Implementation

There are many possible choices of product assortment, even after we restrict our attention to the confections category. However, a large number of these potential assortments are dominated under a wide range of wholesale prices and rebate payments (e.g., replacing Peanut M&M's with the worst-selling product). Taking this into account, we compute the full

[57] The key to successfully implementing this normalization is that the assortment of our hypothetical 'full' machine is a strict superset of any possible observed assortment $a_t \subset a^*$, and that our normalization is the same for all a_t. Under the hypothetical 'full' machine with outside good share $s_0(\delta,\theta,a^*)$, the relationship between the number of consumer arrivals and the state space Δx_t is well defined, and merely requires inflating all of the 'inside good' probabilities by $\frac{1}{1-s_0(a^*)}$. It is easier to match the observed arrival rate in the normalized state space than in the original state space.

[58] The ergodic distribution of x (the number of likely consumers to have arrived since the vending machine was last restocked) as a function of the restocking policy (restock after $x \geq e$ likely consumer arrivals) is given by the solution Γ to the system of equations $\Gamma = \Gamma\tilde{P}(e)$. Likewise, the matrix inverse from eqn (11), $(I - \beta\tilde{P}(e))^{-1}$ represents the present discounted value of a stream of payments. Both of these quantities depend only effort through the post-decision transition matrix $\tilde{P}(e)$). The last term in equation 11, $\hat{u}^R(x,a)$, is our simulated cumulative payoff function and depends only on the assortment (and the state variable x).

payoffs at (a,e) for each agent for 15 possible assortments. Each of the 15 possible assortments includes Mark Vend's five most commonly-stocked chocolate confections products: four Mars products (Snickers, Peanut M&Ms, Twix, and Plain M&Ms), and Nestle's Raisinets. The retailer is always worse off if he replaces any of these five products with a different product. We then allow the retailer to choose any pair of products for the final two slots in the confections category from a set of six products. The six products we consider include two Mars products (Milky Way and Three Musketeers), two Hershey products (Reese's Peanut Butter Cup and PayDay), and two Nestle products (Butterfinger and Crunch).[59] Although we compute the full model for all 15 possible assortments, only three end up being pay-off relevant: (M,M) – 3 Musketeers and MilkyWay, (H,M) – 3 Musketeers and Reese's Peanut Butter Cups, and (H,H) – Reese's Peanut Butter Cup and PayDay.

Finally, Mark Vend's assortment decision is discrete (either a product is on the shelf of the vending machine or it isn't), and our effort decision is discrete (we are restricted to restocking after an integer number of likely consumer arrivals). Thus, we can, and do, enumerate the payoffs of all of the agents at all of the possible assortments a and effort levels e. We assume that the retailer chooses an assortment a and effort level e to maximize his profits, and that he takes wholesale prices and rebate contracts as given. Note that there is no randomness in $\pi(a,e)$, because we evaluate $\pi(a,e)$ using the ergodic distribution of the post-decision transition matrix. This makes it easy to compare across assortments and effort levels.

5 Results

Two features of our data are important to understand for interpreting our results. First, we observe retail prices (fixed at 75 cents for confection products) and wholesale prices that vary across manufacturer, but we do not observe manufacturer costs of production or try to recover them from first order conditions. Thus, in most of our results we report upstream firm revenues, rather than variable profit.[60] Second, although our consumer choice model identifies an ordinal ranking of product assortments for consumers, it does not identify a monetary measure of consumer welfare, because we do not estimate a price coefficient. In

[59] In practice, we consider a larger set of potential products including Mars's Skittles and Mars's Starbust, but those are always dominated by Three Musketeers and MilkyWay. (Skittles and Starburst tend to be more popular with younger customers and less popular with our white-collar professional workers). For some other products, we do not have sufficient information to consider them in our counterfactual analysis. For example, Hershey's with Almonds is popular nationally, but is never available in our data.

[60] This implies that our results provide an upper bound on the gap between the retailer optimal effort level e^R and the vertically-integrated optimal level e^{VI}.

order to convert consumer surplus into dollars, we perform a calibration exercise in which we assume that the median own-price elasticity is -2. We view this as a relatively inelastic estimate of elasticity, implying that our consumer surplus calculations are likely to capture an 'upper bound' on the potential efficiency effects of the AUD.[61]

5.1 Foreclosure with a Fixed Effort Policy

In order to examine foreclosure, holding the effort policy fixed, we parallel our theoretical model from sections 2.1 and 2.2. Our objective is to determine whether foreclosure is possible, and whether foreclosure leads to an assortment that maximizes industry profits. Later, we examine the size of the potential efficiency gains that arise from additional retailer effort. We hold fixed observed wholesale prices and the observed rebate discount λ.

We assume that the retailer chooses an e^R effort policy (from equation (3)).[62] We compute $\pi^R(a,e)$, $\pi^M(a,e)$, and $\pi^H(a,e)$ for the (R)etailer, the dominant firm (M)ars, and the competitor (H)ershey under three assortments, $\{(H,H),(H,M),(M,M)\}$, which we report in table 9.[63] The effort policies are similar across product assortments, and imply that the retailer restocks after $257-261$ likely consumers. As in equation (1), $\pi^R(H,H) > \pi^R(H,M) > \pi^R(M,M)$ or $(36{,}656 > 36{,}394 > 36{,}086)$. Thus, absent any transfers, the retailer prefers to stock (H,H) in the final two slots.

The first column of the second pane of table 9 considers a transfer payment, conditional on replacing Payday (a Hershey product) with Mars' 3 Musketeers (i.e., moving from $(H,H) \rightarrow (H,M)$, partially foreclosing H.[64] This change in assortment reduces retailer profits ($\Delta\pi^R = -262$), but increases the profits of the dominant firm ($\Delta\pi^M = 1{,}657$). The bilateral gains to R and M ($\Delta\pi^R + \Delta\pi^M = 1{,}395$) exceed the losses to the rival ($\Delta\pi^H = -868$). Thus, even if H gives up all of its lost profit, it can not avoid being (partially) foreclosed, and condition (B3) is satisfied. A feasible transfer (by conditions (B1) and (B2)), requires a payment $T \in [262, 1{,}657]$. Therefore, all three conditions (B1), (B2), and (B3), are satisfied, and there exists a feasible transfer that increases total producer surplus (by \$501 units after including Nestle's profits), and also increases consumer surplus by \$261 (assuming a median own-price elasticity of $\epsilon = -2$). Thus, partial foreclosure (H,M) is possible, and increases producer and consumer surplus relative to an (H,H) assortment. However, the value of the

[61] We provide additional details on the calibration exercise, as well as robustness to elasticities of -1 and -4 in Appendix A.7.

[62] We obtain qualitatively similar results across assortments if we assume that the retailer chooses a different effort policy throughout.

[63] We also compute, but do not report separately, Nestle's revenues from sales of Raisinets.

[64] Standard errors are computed based on the procedure outlined in Appendix A.6.

rebate at the observed λ ($\lambda\pi^M(H,M) = 1{,}882$) exceeds the gains to Mars ($\Delta\pi^M = 1{,}657$). Thus, Mars pays more to partially foreclose Hershey than it expects to gain from partial foreclosure. This cannot be an equilibrium outcome.

The second column of the second pane of table 9 starts from (H,M) and considers a move to (M,M). Now Reese's Peanut Butter Cup is replaced by MilkyWay and Hershey is fully foreclosed. Again, the retailer gives up some profit absent the rebate payment ($\Delta\pi^R = -308$), the dominant firm gains ($\Delta\pi^M = 1{,}338$), and bilateral surplus increases ($\Delta\pi^M + \Delta\pi^R = 1{,}030$). However, the gain in bilateral surplus is smaller than the losses to the rival ($\Delta\pi^H = -1{,}299$). This means that (C3) is violated, and (C4) holds instead. Moving from partial foreclosure (H,M) to full foreclosure (M,M) reduces total producer surplus ($\Delta PS = -272$) and consumer surplus (assuming $\epsilon = -2$, $\Delta CS = -110$). Overall social surplus declines ($\Delta SS = -383$). A feasible transfer payment (satisfying (C1) and (C2)) limits the transfer payment $T \in [308, 1338]$. At the observed λ, the value of the additional rebate (assuming the retailer has already moved from (H,H) to (H,M)) would be only 214, which implies that the retailer would reject the rebate and continue to stock an (H,M) assortment, rather than fully foreclosing Hershey.

The final column of the second pane of table 9 considers an assortment change from (H,H) to (M,M) (full foreclosure of both Hershey's products). This is simply the sum of the change in profits from moving first from $(H,H) \to (H,M)$ and then from $(H,M) \to (M,M)$. Again we see that the retailer's variable profit in the absence of the rebate payment decreases ($\Delta\pi^R = -570$), and Mars' revenue increases ($\Delta\pi^M = 2{,}995$), and bilateral surplus rises ($\Delta\pi^M + \Delta\pi^R = 2{,}425$). Meanwhile, Hershey's losses ($\Delta\pi^H = -2{,}167$) are smaller than the gain in bilateral surplus to M and R, implying that (A3) holds. Overall, we see a net gain in producer surplus ($\Delta PS = 229$) and consumer surplus ($\Delta CS = 150$ at $\epsilon = -2$) when the contract induces a change from (H,H) to (M,M). The set of feasible transfers (as defined by (A1) and (A2)) is $T \in [570, 2{,}995]$, which includes our observed rebate payment ($\lambda\pi^M(M,M) = 2{,}096$).

There are a few implications of these findings. The first is that (A1)-(A3), (B1)-(B3), (C1), (C2), and (C4) hold. Therefore, theorem 2 tells us that (M,M) is an equilibrium assortment even though industry profits $\pi^I = \pi^M + \pi^H + \pi^R$ and producer surplus are higher under partial foreclosure (at (H,M)) than they are under full foreclosure (at (M,M)). The second is that, given the observed size of the rebate λ, the rebate is only individually rational for Mars to offer if Mars believes it will cause the retailer to switch from $(H,H) \to (M,M)$. If Mars believes that it would induce a switch only from $(H,H) \to (H,M)$, then the rebate would be too generous. Likewise, if Mars believes that, absent the rebate, the retailer would have stocked (H,M), the rebate would not be generous enough to induce the retailer to switch from $(H,M) \rightarrow (M,M)$.

5.2 Role of the Threshold

These results are meant to parallel those in section 2.3. We explore how the rebate threshold $\overline{\pi}^M$ affects the retailer's choice of assortment and effort, assuming that wholesale prices and the rebate discount λ are fixed at their observed values. Figure 2 plots two curves. Each curve represents the profits of the retailer after receiving the rebate (i.e., $\pi^R(a,e)+\lambda\pi^M(a,e)$). The horizontal axis reports revenue of the dominant firm, π^M. The left curve represents the retailer's profits with an (H,M) assortment. The right curve represents the retailer's profits with a (M,M) assortment. As we move across each curve from left to right, the retailer's effort level is increasing (the policy e is declining). At the peak of each curve is a dot, which represents the $e^R(a)$ level of effort. For reference, we also plot $e^{VI}(a)$ and $e^{SOC}(a)$ for each curve. For any value of $\overline{\pi}^M$ to the left of $e^R((H,M))$, the retailer chooses $e^R((H,M))$. As $\overline{\pi}^M$ increases beyond $\pi^M(e^R(H,M))$, the retailer exerts additional effort in order to meet the rebate threshold. At a 'high enough' threshold, the retailer finds it preferable to foreclose the competitor, rather than exert additional effort. At this threshold, he changes his assortment and jumps to $e^R(M,M)$. We denote this critical threshold with dashed lines. A similar pattern happens with the (M,M) assortment. Mars can induce effort beyond $e^R(M,M)$. However, beyong the point at which Mars products are always available, effort no longer increases π^M. From this point on, no amount of additional effort makes it possible for the retailer to obtain the rebate, and he reverts to (H,H) and $e^{NR}(H,H)$ (not shown in the figure).

We solve the retailer's choice of effort for all possible threshold values $\overline{\pi}^M$ and report the critical threshold values in table 10. The intuition follows exactly the intuition from figure 2. We find that, given no threshold at all, the discount payment λ induces the retailer to switch from (H,H) to (H,M) with an effort policy of $e^R(H,M)$. The retailer stays at $e^R(H,M)$ until $\overline{\pi}^M \geq 11{,}763$. For $\overline{\pi}^M \in [11{,}763, 11{,}912]$ the retailer's choice of effort is dictated by the threshold constraint (including the vertically-integrated level of effort, but not the socially-optimal level of effort). Above this point, for $\overline{\pi}^M \in [11{,}912, 13{,}101]$, the retailer switches to an (M,M) assortment with effort policy $e^R(M,M)$, and fully forecloses Hershey. As the threshold increases further, the retailer increases his effort in order to satisfy the threshold (including the vertically-integrated e^{VI} and socially-optimal e^{SOC} effort levels), until he reaches $\overline{\pi}^M = 13{,}320$. For any threshold higher than this, no amount of additional effort makes the rebate achievable, and the retailer reverts to an $(H{,}H)$ assortment with effort level $e^{NR}(H{,}H)$.

5.3 Effort and Potential Efficiency Gains

Table 11 reports the effort policies from equation (3) for all three assortments.[65] For the socially-optimal effort policy, we report three calculations, which differ based on our assumption about consumers' median own-price elasticity of demand, ranging from inelastic ($\epsilon = -1$) to more elastic ($\epsilon = -4$).

In general, the effort levels for a given policy are relatively similar across the three product assortments. We discuss results for the case in which Hershey is fully foreclosed $a = (M{,}M)$. Absent the rebate, the retailer chooses to restock after every 264 consumers (i.e., an effort level of $e^{NR} = 264$). The discount aspect of the rebate, λ, reduces the effective wholesale price to the retailer, and leads him to increase his effort ($e^R = 259$). Maximizing bilateral surplus of $M + R$ leads to a higher effort choice ($e^{VI} = 243$). Maximizing surplus of the entire industry (the retailer plus the three confections manufacturers) reduces effort slightly, relative to e^{VI}, so that $e^{IND} = 244$. We demonstrate why this happens in figure 3, which plots $u'(x)$ (the incremental variable profit per consumer – ignoring the retailer's restocking cost) for the $(H{,}M)$ assortment, for the retailer and each of the three manufacturers.[66] For effort levels $e \in [200{,}400]$, lower effort monotonically decreases the incremental variable per-consumer profits of both Mars and the retailer, but increases the profits of the rivals Hershey and Nestle. This happens because sales of Hershey and Nestle products increase when the best-selling Mars products (Peanut M&Ms and Snickers) stock out. We previously documented this phenomenon with our experiment in table 6.

Finally, table 11 considers the effort policies that account for consumer surplus. These policies vary with the assumed median own-price elasticity of demand, but generally imply a higher level of effort ($e^{SOC} = \{235{,}229{,}222\}$ for $\epsilon = \{-4, -2, -1\}$ respectively). In Appendices A.7 and A.8, we show that more inelastic demand is akin to the social planner placing more weight on consumer surplus.

In table 12 we consider the potential efficiency gains induced by the AUD rebate, holding assortment fixed. We focus on the vertically-integrated (e^{VI}) and socially-optimal (e^{SOC}),

[65] These calculations implicitly ignore the role of the threshold.

[66] The concavity of $u(x)$ is guaranteed because $u'(x)$ is always (weakly) positive (i.e., it is bounded by the case where all products stock out) and is (weakly) decreasing (i.e., additional products continue to stock-out whenever the wait time between restocking visits is increased).

effort levels.[67] Under an (M,M) assortment, switching from the no-rebate retailer effort policy (e^{NR}) to the vertically-integrated optimal effort policy (e^{VI}) increases the restocking frequency by 7.95%. We use the e^{NR} effort level rather than the e^R effort level as our baseline in order to capture the maximum potential efficiency gains from the rebate contract.[68] Higher effort is costly to the retailer ($\Delta\pi^M = -55$) and beneficial to Mars ($\Delta\pi^M = 128$), leading to a net gain in producer surplus ($\Delta PS = 63$) once we include competing manufacturers. Most of the gains to effort accrue to consumers ($\Delta CS = 192$). The net gain in social surplus is positive ($\Delta SS = 255$). The socially-optimal effort policy leads to an even larger increase in the frequency of restocking (13.26% for the (M,M) assortment), smaller gains in producer surplus ($\Delta PS = 17$), and larger gains in consumer surplus ($\Delta CS = 284$). The net gain in social surplus is again positive ($\Delta SS = 301$).

In table 13 we compare welfare calculations under the (M,M) assortment with the e^R, e^{VI}, and e^{SOC} effort levels against two potential baselines. The first baseline is the (H,H) assortment under the e^{NR} effort level. This mimics what the retailer would choose if the AUD contract were to disappear but wholesale prices were to remain fixed. The second baseline is the (H,M) assortment with the e^{NR} effort level. This is the assortment that would be chosen by the social planner, but without efficiency gains from lower wholesale prices or the rebate threshold.

Starting from a base assortment of (H,H), the AUD contract that leads to an (M,M) assortment increases both consumer and producer surplus. Even when the threshold does not induce additional effort beyond e^R, social surplus increases ($\Delta SS = 477$). Producer and consumer surplus further increase under the vertically-integrated effort level ($\Delta SS = 654$). The socially-optimal effort policy increases consumer surplus at the expense of producer surplus, for a net gain of $\Delta SS = 700$.

When we compare the (M,M) assortment to the socially-optimal assortment (H,M), we find that if the AUD threshold $\overline{\pi}^M$ is set at a high enough level to obtain exclusion, but not high enough to incentivize additional effort beyond e^R, then producers and consumers are both worse off than they would be absent the rebate ($\Delta PS = -239$ and $\Delta CS = -49$). If, however, the threshold is set to obtain the vertically-integrated effort level, e^{VI}, producer surplus declines ($\Delta PS = -203$), but consumers benefit ($\Delta CS = 92$). At a median own-price elasticity of $\epsilon = -2$, this implies that the net effect is negative ($\Delta SS = -111$). Even at the

[67] We use a median own-price demand elasticity of $\epsilon = -2$ to calibrate consumer surplus in table 12.

[68] Most additional retail effort comes through Mars' choice of the threshold, rather than its choice of λ, because when we compare the e^{VI} restocking frequency to the e^R optimal (which includes the lower wholesale price), the increase in frequency is still 6.18%.

socially-optimal effort level, e^{SOC}, the loss in producer surplus ($\Delta PS = -250$) exceeds the gains to consumers from the higher effort ($\Delta CS = 185$).

5.4 Role of the Discount and Competitive Response

Thus far, we have addressed the question of whether Hershey could avoid foreclosure by comparing the bilateral gains of Mars and the retailer to Hershey's losses using conditions like (A3). The notion behind this comparison is that if Hershey does not have enough surplus to transfer to the retailer, any attempts to prevent foreclosure through wholesale price cuts will be futile. Table 14 explores the mechanics of how this might play out.

Holding the wholesale prices of Mars and Nestle (w_m,w_n), and λ, fixed, we let Hershey adjust its price from $w_h \rightarrow w_h'$, where we define w_h' as the wholesale price that makes the retailer indifferent between accepting the rebate payment and foreclosing Hershey, and purchasing the Hershey products at the reduced wholesale price. We compute the critical values of w_h' and report them in table 14. If the alternative to the (M,M) assortment with a rebate payment is an assortment of (H,H) with effort level e^{NR}, we find that Hershey would need to cut its wholesale price to $12.83 - 15.35$ cents, depending on how much additional effort is induced by the rebate threshold (i.e., e^R, e^{VI}, or e^{SOC}). This is remarkably close to our best estimate of the marginal cost of production, which industry sources identify as around 15 cents per candy bar. Thus, at the existing level of the rebate λ, Hershey would likely need to sell at a loss in order to avoid foreclosure. This suggests that Hershey may not be able to avoid foreclosure by outbidding Mars for placement in the retail assortment (even without a potential Mars response to increase λ).[69]

Conversely, we consider whether Mars could reduce its discount and still foreclose its rival. We assume that each manufacturer's production cost is \$0.15 per unit, so that Hershey's best offer to the retailer is a wholesale price of $w_h' = c_h = \$0.15$. We solve for the discount λ' that makes the retailer indifferent between an (H,H) assortment at a wholesale price of w_h' and an (M,M) assortment with a rebate payment. This allows us to calculate whether (and by how much) Mars could reduce its rebate payment and still foreclose Hershey. The precise estimates depend on the effort level induced by the rebate threshold. Under an effort policy of e^R, Mars can reduce the rebate amount by 5.27%, while at e^{VI}, Mars can reduce the rebate by only 3.53%. At the socially-optimal effort, the rebate is not generous enough. Assuming that the \$0.15 production cost estimate is reasonable, this gives an indication that the terms of Mars's current rebate program are well designed.

[69]Formally, (A2) provides Mars' IR constraint for an upper bound on λ.

The right panel of table 14 conducts the same exercise, but assumes that the retailer would choose the (H,M) assortment in the absence of Mars' rebate. Under this scenario, the rebate is much too generous and could be reduced by 38.18% to 44.79% while still foreclosing Hershey. Relatedly, holding λ fixed, Hershey would need to set a negative wholesale price (i.e., pay the retailer to sell its products). This highlights the fact that the rebate terms are only sensible as a device to make the retailer switch from $(H,H) \rightarrow (M,M)$.[70]

5.5 Comparison to Uniform Wholesale Pricing by Mars

In lieu of an AUD, Mars could charge a lower wholesale price without conditioning on a threshold $\overline{\pi}^M$. Table 15 presents results for a uniform wholesale price by Mars. We hold fixed the wholesale prices of Hershey and Nestle (w_h,w_n), and compute a new optimal wholesale price for M, w'_m. The resulting set of wholesale prices (w'_m,w_h,w_n) does not constitute an equilibrium (because (w_h,w_n) are not allowed to adjust). Therefore, the exercise is meant as tool to understand how the AUD reduces the price of foreclosure to the dominant firm, rather than reflecting what would happen to equilibrium prices in the absence of an AUD by Mars.

The main result is that Mars' wholesale price is lower than the post-discount wholesale price under the AUD. Effectively, Mars pays more for foreclosure without the threshold. We quantify exactly how much more by comparing Mars' uniform wholesale price to an AUD that forecloses under two different effort levels: e^R and e^{VI}. Mars' profit (after rebates), $(1-\lambda)\pi^M$, falls from \$11,005 to \$10,094, for a loss of \$911. The retailer's profit increases by a similar amount (\$921, or \$39,103 - \$38,182). The gains to the retailer are slightly larger if the threshold under the AUD had been used to implement the vertically-integrated effort level.

We plot the best response of Mars to the observed (w_h,w_n) prices in figure 4. We do not consider an equilibrium in which all three upstream firms simultaneously set wholesale prices (w_m,w_h,w_n). The challenge for modeling this is that no Nash equilibrium exists in pure strategies because the retailer's assortment decision is discrete; only a mixed-strategy Nash equilibrium exists.[71]

[70] It should also be clear that adjusting the baseline from (H,H) to (H,M) means that the current rebate violates Mars's IR constraint (B2) as noted in table 9.

[71] The non-existence of pure-strategy equilibria is well documented in the theoretical literature (e.g., see recent work by Jeon and Menicucci (2012)), and derives from the fact that agents' best-response functions are discontinuous, and need not cross. The mixed-strategy Nash equilibrium of the uniform wholesale pricing game is not easily interpretable in our context, and is beyond the scope of our analysis of Mars' AUD contract.

5.6 Implications for Mergers

Vending is one of many industries for which retail prices are often fixed across similar products and under different vertical arrangements. Indeed, there are many industries for which the primary strategic variable is not retail price, but rather a slotting fee or other transfer payment between vertically-separated firms. Thus, our ability to evaluate the impact of a potential upstream merger may turn on how the merger affects payments between firms in the vertical channel. We consider the impact of three potential mergers (Mars-Hershey, Mars-Nestle, and Hershey-Nestle) on the AUD terms offered to the retailer by Mars. Given the degree of concentration in the confections industry, antitrust authorities would likely investigate proposed mergers, especially mergers involving Mars.[72]

Table 16 measures how competing manufacturers might respond to an upstream merger. The first column duplicates the second column of table 14 as a baseline. In the second column, we examine a potential Mars-Hershey merger. We assume that after the merger, the Hershey product (Reeses Peanut Butter Cup) is priced at the Mars wholesale price and included in Mars' rebate contract. The merged (Mars-Hershey) firm is now happy for consumers to substitute to Reese's Peanut Butter Cups, and the AUD is able to achieve the industry-optimal (and socially-optimal) product assortment of (H,M).[73] The merged firm faces competition from Nestle (Nestle Crunch and Butterfinger), which charges lower wholesale prices but sells less-popular products.[74] In the absence of an AUD, the retailer maximizes his profit by stocking the two Nestle products, but the AUD induces the retailer to choose an (H,M) assortment and an e^{VI} effort policy (evaluated at the observed discount λ and wholesale prices).[75] Although not reported in table 16, the AUD also maximizes social surplus by inducing the socially-optimal (H,M) assortment and a high effort level.[76]

We consider the possibility that Nestle may be able to cut its price in order to avoid having Butterfinger and Nestle Crunch foreclosed. Following the same exercise that was performed in table 14, we find that Nestle would need to charge a negative wholesale price to the retailer in order to induce him to stock the less-popular Nestle products (similar to condition (A3)). Knowing that the Nestle productss provide weak discipline for the merged Mars-Hershey firm, we next examine whether the merged Mars-Hershey firm can reduce the

[72]For a related analysis of diversion ratios in this market, see Conlon and Mortimer (2013b).

[73]We assume that the AUD retains λ at the pre-existing level, and sets $\overline{\pi^M} = \pi^M(e^{VI}(H,M))$ to induce the vertically-integrated optimal level of effort.

[74]We use Nestle's observed wholesale price when computing changes in profits and producer surplus.

[75]Table 16 reports changes in variable profit for each agent, but not levels. For the full details of post-merger profits (or revenues for manufacturers) at all $\pi(a,e)$, please see Appendix A.9.

[76]Producer surplus and consumer utility for each potential merger are also reported in Appendix A.9.

generosity of its rebate. Pre-merger, we found that the rebate was 3.53% too generous; after the merger it is 42.3% too generous. This implies that the market is unlikely to obtain the post-merger outcome in which the retailer stocks the socially-optimal assortment, because the discipline imposed by Nestle's products is likely too weak to keep the current AUD terms in force.

We perform a similar exercise in the third column, in which we allow Mars and Nestle to merge. The main difference now is that the merged firm internalizes the profits of Nestle's Raisinets, and is able to include the profit from Raisinets in the rebate. This again provides incentives for the merged firm to reduce the generosity of the rebate (by 12.67%).[77] Finally, we examine a Hershey-Nestle merger in the final column. Giving Hershey access to the profits of Raisinets does very little, because Raisinets is not in danger of being foreclosed. This exercise closely resembles our baseline (No Merger) scenario.

Throughout the paper, we report the variable profits for the retailer; it is likely that his overall operating profits, after accounting for administrative and overhead costs, are substantially lower. In the anti-trust case involving Intel's loyalty contract with Dell, Intel's rebate program was reported to account for more than one quarter of Dell's operating profits. Based on communication with industry participants, we think that the Mars rebate may account for an even larger fraction of operating profits in the vending industry. This means that a 42% rebate reduction (implied by the hypothetical Mars-Hershey merger) may represent a substantial fraction of the overall operating profits of the retailer.

6 Conclusion

Using a new proprietary dataset that includes exogenous variation in product availability, we provide empirical evidence regarding the potential efficiency and foreclosure aspects of an AUD contract. Similar vertical rebate arrangements have been at the center of several recent large antitrust settlements, and have attracted the attention of competition authorities in many jurisdictions.

In order to understand the relative size of the potential efficiency and foreclosure effects of the contract, our framework incorporates endogenous retailer effort and product assortment decisions. A model of consumer choice allows us to characterize the downstream substitutability of competing products, and combining this with a model of retailer effort allows us to estimate the impact of downstream effort across upstream and downstream firms.

[77] The only difference comes through the effort channel, as Mars prefers less effort once it internalizes substitution to Raisinets.

Identification of both the consumer choice and retailer-effort models benefits from exogenous variation in product availability made possible through a field experiment. We show that the vertical rebate we observe has the potential to increase effort provision by roughly 9-11%, and that the benefit of this additional effort is mostly captured by consumers. The rebate also enables the dominant firm, Mars, to foreclose Hershey by leveraging its profits from dominant products (such as Snickers and Peanut M&Ms), to obtain shelf-space for products such as Milky Way.

We find that at the prevailing wholesale prices, this foreclosure enhances the profitability of the overall industry and improves social surplus, but does not lead to a product assortment that maximizes industry profits. We note that in the absence of the vertical rebate, manufacturers may charge different wholesale prices. In a limited comparison of Mars' optimal linear wholesale prices to the AUD contract, we find that the primary difference between Mars' AUD and linear wholesale pricing is the allocation of profits between the dominant upstream firm and the retailer. The differential impact on social welfare is small, and depends on how the dominant firm sets the quantity threshold in the AUD. Finally, we explore the potential impact of three hypothetical upstream mergers on the likely terms of the AUD contract, holding retail prices fixed. We find that a merger between the two largest upstream firms has the potential to induce the socially-optimal product assortment, but may also lead to a reduction in the rebate payments made to retailers.

In addition to providing a road-map for empirical analyses of vertical rebates, and results on one specific vertical rebate, our detailed data and exogenous variation allow us to contribute to the broader literature on the role of vertical arrangements for mitigating downstream moral hazard and inducing downstream effort provision. Empirical analyses of downstream moral hazard are often limited not only by data availability, but also by the ability to measure effort, and our setting proves a relatively clean laboratory for measuring the effects of downstream effort.

References

Aghion, P., and P. Bolton (1987): "Contracts as Barriers to Entry," *American Economic Review*, 77(3), 388–401.

Asker, J. (2016): "Diagnosing Foreclosure due to Exclusive Dealing," *Journal of Industrial Economics*, 64(3), 375–410.

Asker, J., and H. Bar-Isaac (2014): "Raising Retailers' Profits: On Vertical Practices and the Exclusion of Rivals," *American Economic Review*, 104(2), 672–686.

Baker, G. P., and T. N. Hubbard (2003): "Make Versus Buy in Trucking: Asset Ownership, Job Design, and Information," *American Economic Review*, 93(3), 551–572.

Bernheim, B. D., and M. Whinston (1998): "Exclusive Dealing," *Journal of Political Economy*, 106(1), 64–103.

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

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Bork, R. (1978): *The Antitrust Paradox*. New York: Free Press.

Brickley, J., and F. Dark (1987): "The Choice of Organizational Form: The Case of Franchising," *Journal of Financial Economics*, 18(2), 401–420.

Chao, Y., and G. Tan (2014): "All-Units Discounts as a Partial Foreclosure Device," Working Paper.

Chen, C.-W. (2014): "Estimating the Foreclosure Effect of Exclusive Dealing: Evidence from the Entry of Specialty Beer Producers," .

Chen, Z., and G. Shaffer (2014): "Naked Exclusion with Minimum-Share Requirements," *Rand Journal of Economics*, 45(1), 64–91.

Chipty, T. (2001): "Vertical Inegration, Market Foreclosure, and Consumer Welfare in the Cable Television Industry," *American Economic Review*, 91(3), 428–53.

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

Conlon, C. T., and J. H. Mortimer (2010): "Effects of Product Availability: Experimental Evidence," NBER Working Paper No. 16506.

——— (2013b): "An Experimental Approach to Merger Evaluation," NBER Working Paper No. 19703.

Crawford, G., R. Lee, M. Whinston, and A. Yurukoglu (2015): "The Welfare Effects of Vertical Integration in Multichannel Television Markets," NBER Working Paper No. w21832.

Deneckere, R., H. Marvel, and J. Peck (1996): "Demand Uncertainty Inventories and Resale Price Maintenance," *Quarterly Journal of Economics*, 111(3), 885–913.

——— (1997): "Demand Uncertainty and Price Maintenance: Markdowns as Destructive Competition," *American Economic Review*, 87(4), 619–641.

Elhauge, E., and A. Wickelgren (2012): "Anticompetitive Market Division through Loyalty Discounts without Buyer Commitment," Harvard Discussion Paper No. 723; U of Texas Law, Law and Econ Research Paper No. 239.

——— (2014): "Robust Exclusion and Market Division Through Loyalty Discounts with Buyer Commitment," Working Paper.

Figueroa, N., E. Ide, and J.-P. Montero (2016): "Discounts as a Barrier to Entry," *American Economic Review*, 106(7).

Fumagalli, C., and M. Motta (2006): "Exclusive Dealing and Entry, when Buyers Compete," *American Economic Review*, 96(3), 785–795.

Genchev, B., and J. H. Mortimer (forthcoming): "Empirical Evidence on Conditional Pricing Practices: A Review," *Anti-trust Law Journal*.

Gil, R. (2007): "Make-or-buy in Movies: Integration and Ex-post Renegotiation," *International Journal of Industrial Organization*, 25(4), 643–655.

Ho, J., K. Ho, and J. H. Mortimer (2012a): "Analyzing the Welfare Impacts of Full-line Forcing Contracts," *Journal of Industrial Economics*, 60(3), 468–498.

——— (2012b): "The Use of Full-line Forcing Contracts in the Video Rental Industry," *American Economic Review*, 102(2), 686–719.

Hotz, V. J., and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60(3), 497–529.

Hubbard, T. N. (1998): "An empirical examination of moral hazard in the vehicle inspection market," *Rand Journal of Economics*, 29(2), 406–426.

Jeon, D.-S., and D. Menicucci (2012): "Bundling and Competition for Slots," *American Economic Review*, 102(5), 1957–85.

Kapor, A. (2008): "Queueing and Novelty Effects in Consumer Demand," Harvard Archives.

Klein, B., and K. M. Murphy (1988): "Vertical Restraints as Contract Enforcement Mechanisms," *Journal of Law and Economics*, 31(2), 265–297.

Kolay, S., G. Shaffer, and J. A. Ordover (2004): "All-Units Discounts in Retail Contracts," *Journal of Economics and Managment Strategy*, 13(3), 429–459.

Lafontaine, F. (1992): "Agency Theory and Franchising: Some Empirical Results," *RAND Journal of Economics*, 23(2), 263–283.

Lafontaine, F., and M. Slade (2007): "Vertical Integration and Firm Boundaries: The Evidence," *Journal of Economic Literature*, 45(3), 629–685.

——— (2008): "Exclusive Contracts and Vertical Restraints: Empirical Evidence and Public Policy," in *Handbook of Antitrust Economics*, ed. by P. Buccirossi, pp. 391–414. MIT Press, Cambridge, MA.

Lee, R. S. (2013): "Vertical Integration and Exclusivity in Platform and Two-Sided Markets," *American Economic Review*, 103(7), 2960–3000.

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, pp. 75–96. North-Holland.

O'Brien, D. P. (2013): "All-Units Discounts and Double Moral Hazard," FTC Working Paper No. 316.

Posner, R. (1976): *Antitrust Law: An Economic Perspective*. University of Chicago Press.

Rasmusen, E. B., J. M. Ramseyer, and J. Wiley, John S. (1991): "Naked Exclusion," *The American Economic Review*, 81(5), 1137–1145.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

Sass, T. (2005): "The competitive effects of exclusive dealing: Evidence from the U.S. beer industry," *International Journal of Industrial Organization*, 23(3-4), 203–225.

Segal, I. R., and M. D. Whinston (2000): "Naked Exclusion: Comment," *American Economic Review*, 90(1), 296–309.

Shaffer, G. (1991a): "Capturing Strategic Rent: Full-Line Forcing, Brand Discounts, Aggregate Rebates, and Maximum Resale Price Maintenance," *Journal of Industrial Economics*, 39(5), 557–575.

——— (1991b): "Slotting Allowances and Resale Price Maintenance: A Comparison of Facilitating Practices," *RAND Journal of Economics*, 22(1), 120–135.

Shepard, A. (1993): "Contractual Form, Retail Price, and Asset Characteristics in Gasoline Retailing," *RAND Journal of Economics*, 24(1), 58–77.

Sinkinson, M. (2014): "Pricing and Entry Incentives with Exclusive Contracts: Evidence from Smartphones," Working Paper.

Stokey, N. L., R. Lucas, and E. Prescott (1989): *Recursive Methods for Economic Dynamics*. Harvard University Press.

Sudhir, K., and V. R. Rao (2006): "Do Slotting Allowances Enhance Efficiency or Hinder Competition?," *Journal of Marketing Research*, 43(2), 137–155.

Telser, L. (1960): "Why Should Manufacturers Want Fair Trade?," *Journal of Law and Economics*, 3, 86–105.

Tirole, J. (1988): *The Theory of Industrial Organization*. MIT Press.

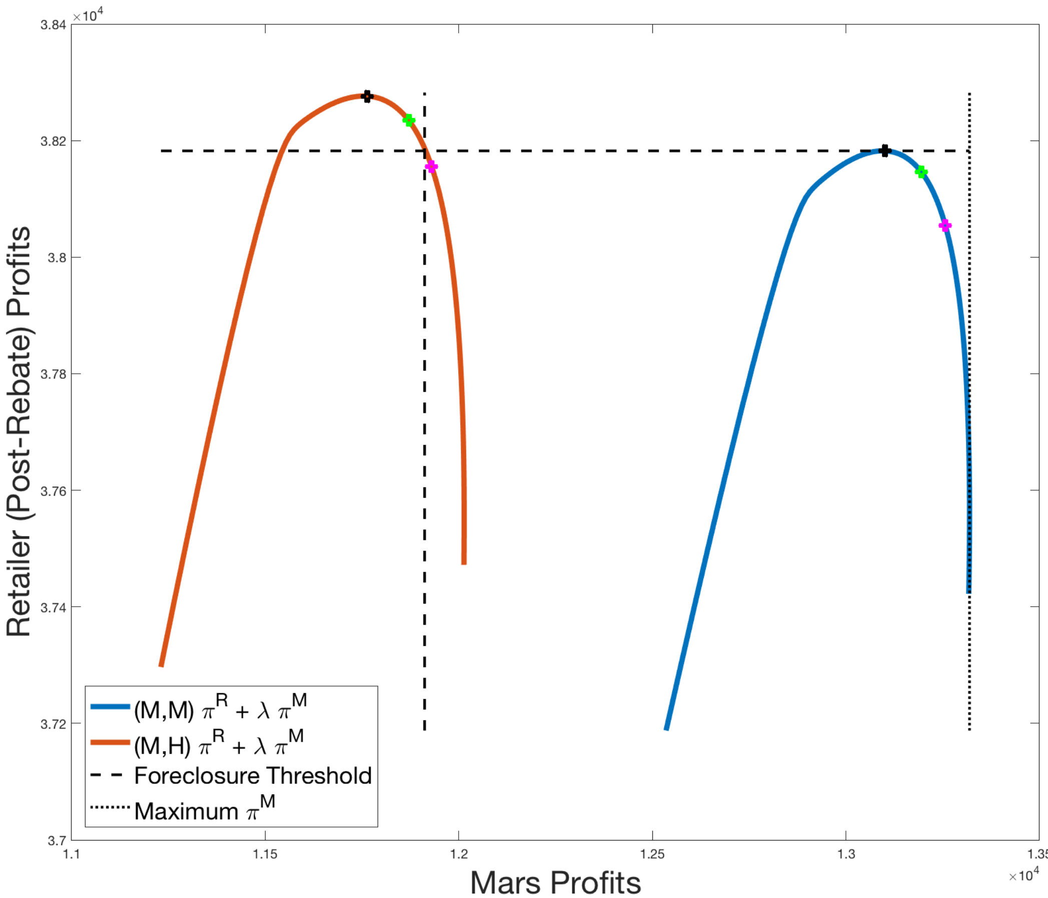
Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

Whinston, M. (1990): "Tying, Foreclosure, and Exclusion," *American Economic Review*, 80(4), 837–59.

Figure 1: Mars Vend Operator Rebate Program

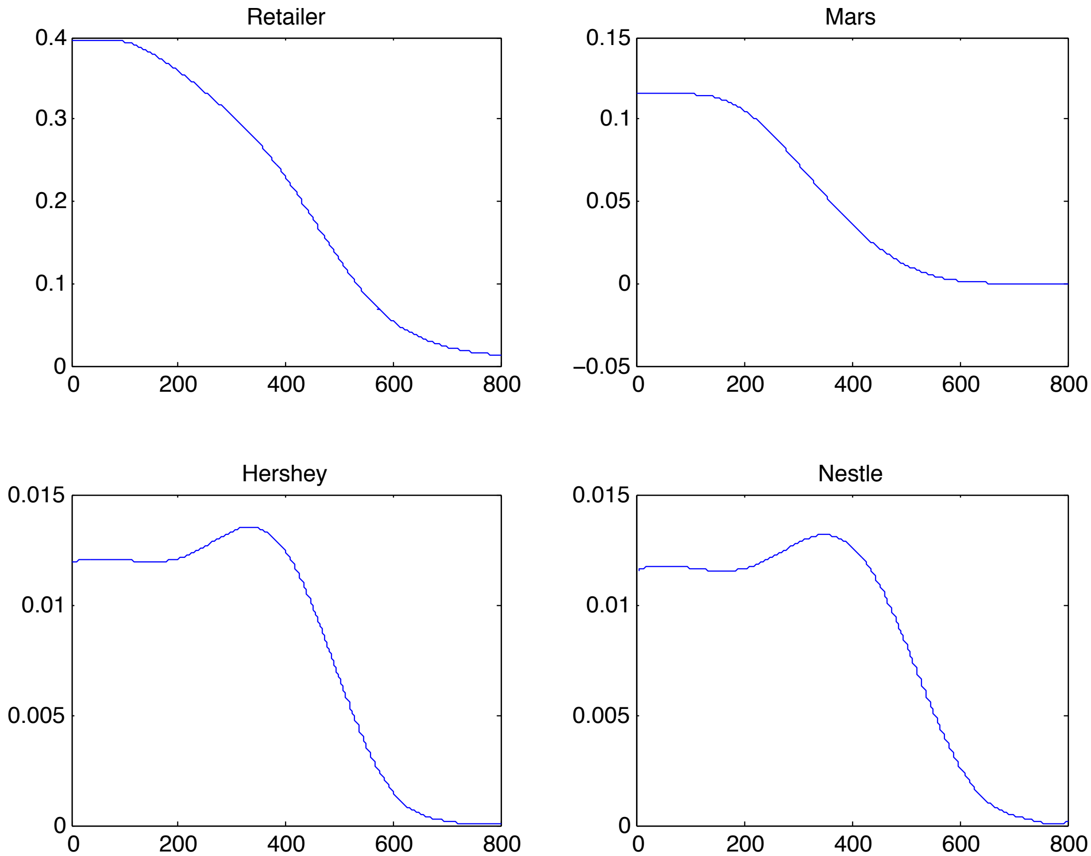
Notes: From '2010 Vend Program' materials, dated December 21, 2009; last accessed on February 2, 2015 at http://vistar.com/KansasCity/Documents/Mars%202010%20Operatopr%20rebate%20program.pdf.

Figure 2: Impact of AUD Quantity Threshold on Retail Assortment Choice



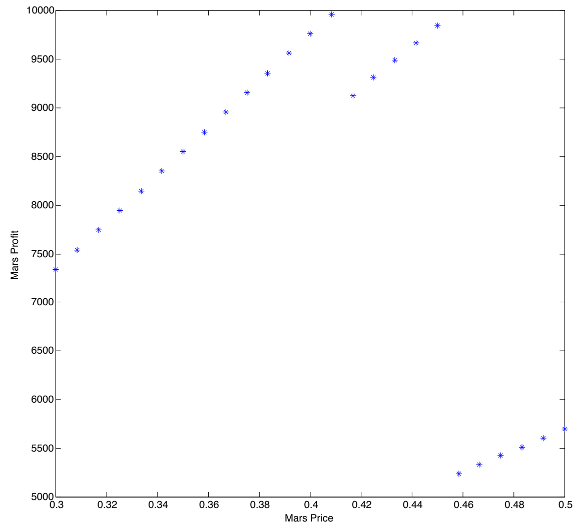
Notes: Figure reports retailer variable profit under two assortment choices ((H,M) on the left and (M,M) on the right), against revenues of Mars products. For a threshold $\overline{\pi}^M \geq 11{,}912$ (noted by the vertical dashed line), the retailer prefers to switch his assortment from (H,M) to (M,M). Three points are marked on each curve. The left-most point on each curve represents an e^R effort policy for the relevant assortment; the point to the right of e^R represents an e^{VI} effort policy, and the right-most point represents an e^{SOC} effort policy.

Figure 3: Profits Per Consumer as a Function of the Restocking Policy



Notes: Each curve reports the profits of the retailer, Mars, Hershey and Nestle as a function of the retailer's restocking policy, using the product assortment in which the retailer stocks 3 Musketeers (Mars) and Reese's Peanut Butter Cups (Hershey) in the final two slots. Specifically, the vertical axes report variable profit per consumer for each of the four firms, and the horizontal axes report the number of expected sales between restocking visits.

Figure 4: Mars Profits as a Function of Price (Linear Pricing)



Notes: Reports Mars' profit at different linear wholesale prices, holding fixed the wholesale prices of Hershey and Nestle. The discontinuities reflect prices at which the retailer drops a Mars product from its assortment.

Table 1: Comparison of National Availability and Shares with Mark Vend

Manufacturer	Product	National: Rank	National: Availability	National: Share	Mark Vend: Availability	Mark Vend: Share	Experimental: Availability	Experimental: Share
Mars	Snickers	1	89	12	87	16.9	97	21.3
Mars	Peanut M&Ms	2	88	10.7	89	16.0	97	22.1
Mars	Twix Bar	3	67	7.7	80	12.6	79	13.0
Hershey	Reeses Peanut Butter Cups	4	72	5.5	71	6.6	45	6.2
Mars	Three Musketeers	5	57	4.3	35	3.1	41	5.2
Mars	Plain M&Ms	6	65	4.2	71	6.6	45	6.2
Mars	Starburst	7	38	3.9	41	3.2	16	1.0
Mars	Skittles	8	43	3.9	65	5.6	79	6.3
Nestle	Butterfinger	9	52	3.2	32	2.1	32	2.6
Hershey	Hershey with Almond	10	39	3	1	0.1	0	0.0
Hershey	PayDay	11	47	2.9	13	1.2	1	0.1
Mars	Milky Way	13	39	1.7	33	2.8	18	1.5
Nestle	Raisinets	>45	N/R	N/R	45	4.0	81	8.7

Notes: National Rank, Availability and Share refers to total US sales for the 12 weeks ending May 14, 2000, reported by Management Science Associates, Inc., at http://www.allaboutvending.com/studies/study2.htm, accessed on June 18, 2014. National figures are not reported for Raisinets because they are outside of the 45 top-ranked products. By manufacturer, the national shares of the top 45 products (from the same source) are: Mars 52.0%, and Hershey 20.5%. For Mark Vend, shares are: Mars 73.6%, and Hershey 15.0% and for our experimental sample Mars 78.3% and Hershey 13.1% (calculations by authors).

Table 2: Assortment Response to Changes in the Threshold

	Achieved Threshold %	Total Vends	Mars Share
2007q1	109.16	1000.00	20.20
2007q2	106.29	1087.45	19.77
2007q3	100.81	1008.57	20.94
2007q4	105.23	1092.49	19.97
2008q1	106.27	1103.42	19.45
2008q2	97.20	1057.32	19.77
2008q3	91.88	1014.13	19.14
2008q4	87.02	1048.26	18.11
2009q1	87.03	1058.54	17.65

Notes: Achieved threshold % reports the ratio of total Mars sales relative to Mars sales in the same quarter one year prior. For quarters 2007q1-2008q1 we believe the target to be 100% with a bonus payment at 105%. For quarters 2008q3-2009q1 we believe the threshold was reduced to 90%.

Table 3: Average Number of Confections Facings Per Machine-Visit

				Mars		Hershey	
	Mars	Hershey	Nestle	Milkyway	3 Musketeer	PB Cup	Payday
2006q1	6.64	1.32	2.05	0.26	0.50	0.19	0.08
2006q2	6.70	1.06	2.02	0.26	0.49	0.15	0.03
2006q3	6.76	0.81	2.02	0.29	0.56	0.03	0.01
2006q4	6.74	0.85	2.00	0.31	0.55	0.01	0.04
2007q1	6.61	1.13	1.58	0.32	0.56	0.00	0.08
2007q2	6.24	1.44	1.17	0.31	0.53	0.00	0.18
2007q3	6.21	1.63	1.08	0.29	0.54	0.01	0.21
2007q4	6.26	1.73	1.03	0.30	0.51	0.15	0.20
2008q1	5.98	2.08	0.97	0.38	0.29	0.51	0.19
2008q2	5.57	2.29	0.93	0.43	0.03	0.66	0.21
2008q3	5.37	2.29	0.91	0.41	0.00	0.63	0.23
2008q4	5.48	2.19	0.89	0.40	0.01	0.62	0.24
2009q1	5.32	1.99	0.83	0.37	0.01	0.62	0.23

Notes: Figures represent the weighted average number of product facings per machine-visit for the entire MarkVend enterprise (117,428 visits). Each machine visit is weighted by overall machine-visit sales to confer more weight on higher-volume machines. This is not a balanced panel, and composition of machine-visits may vary over time for reasons unrelated to assortment decisions. Changes in total facings may be due to: facings by other confections producers, substitution between confections and non-confections products, or changes in visit frequency across different machines.

Table 4: Effort Response to Changes in the Threshold

	Vends Per Visit	Elapsed Days Per Visit
Lower Threshold	8.262***	0.857***
	(0.410)	(0.0690)
Observations	117,428	117,428
R-squared	0.361	0.154
Machine FE	YES	YES
Week of Year FE	YES	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table reports linear regression analysis of 'Vends Per Visit' and 'Elapsed Days Per Visit' on an indicator for the use of a lower AUD threshold by Mars, which occurs beginning in the third quarter of 2008. Results use Mark Vend's entire population of snack vending machines, and include fixed effects for machines and week-of-year. An observation is a service visit at a snack vending machine.

Table 5: Downstream Profit Impact

			Without Rebate			With Rebate		
Exogenous			Difference In:		T-Stat	Difference In:		T-Stat
Removal	Vends	Obs	Margin	Profit	of Diff	Margin	Profit	of Diff
Snickers	-216.82	109	0.39	-56.75	-2.87	0.24	-73.26	-4.33
Peanut M&Ms	-197.58	115	0.78	-10.74	-0.58	0.51	-39.37	-2.48
Double	-282.66	89	1.67	-4.54	-0.27	1.01	-54.87	-3.72

Notes: Calculations by authors, using exogenous product removals from the field experiment.

Table 6: Upstream (Manufacturer) Profits

					% Borne by Mars	
					Without	With
Exogenous Removal	Mars	Hershey	Nestle	Other	Rebate	Rebate
Snickers	-26.37	5.89	19.32	-20.26	31.7%	11.9%
Peanut M&Ms	-68.38	32.76	11.78	-9.36	86.4%	50.2%
Snickers + Peanut M&Ms	-130.81	61.43	20.22	37.10	96.7%	59.5%

Notes: Calculations by authors, using exogenous product removals from the field experiment. The '% Borne by Mars Without Rebate' reports the percentage of the total cost of a product removal that is borne by Mars, without accounting for the rebate payment to the retailer. '% Borne by Mars With Rebate' is equivalently defined.

Table 7: Random Coefficients Choice Model

	Parameter Estimates	
σ_{Salt}	0.506	0.458
	[.006]	[.010]
σ_{Sugar}	0.673	0.645
	[.005]	[.012]
σ_{Peanut}	1.263	1.640
	[.037]	[.028]
# Fixed Effects ξ_t	15,256	2,710
LL	-4,372,750	-4,411,184
BIC	8,973,960	8,863,881
AIC	8,776,165	8,827,939

Notes: The random coefficients estimates correspond to the choice probabilities described in section 4, equation 5. Both specifications include 73 product fixed effects. Total sales are 2,960,315.

Table 8: Products Used in Counterfactual Analyses

'Typical Machine' Stocks:	
Confections:	Salty Snacks:
Peanut M&Ms	Rold Gold Pretzels
Plain M&Ms	Snyders Nibblers
Snickers	Ruffles Cheddar
Twix Caramel	Cheez-It Original
Raisinets	Frito
Cookie:	Dorito Nacho
Strawberry Pop-Tarts	Cheeto
Oat 'n Honey Granola Bar	Smartfood
Grandma's Chocolate Chip Cookie	Sun Chip
Chocolate Chip Famous Amos	Lays Potato Chips
Raspberry Knotts	Baked Lays
Other:	Munchos Potato Chips
Ritz Bits	Hot Stuff Jays
Ruger Vanilla Wafer	
Kar Sweet & Salty Mix	
Farley's Mixed Fruit Snacks	
Planter's Salted Peanuts	
Zoo Animal Cracker Austin	

Notes: These products form the base set of products for the 'typical machine' used in the counterfactual exercises. For each counterfactual exercise, two additional products are added to the confections category, which vary with the product assortment selected for analysis.

Table 9: Assortment Decisions with Fixed Effort

	(H,H)	(H,M)	(M,M)
e^R	257	261	259
π^R	36,656	36,394	36,086
$\lambda\pi^M$	1,617	1,882	2,096
π^M	10,106	11,763	13,101
π^H	2,167	1,299	0
$\pi^R + \pi^M$	46,762	48,157	49,187
$\pi^R + \pi^M + \pi^H$	48,929	49,456	49,187
from	(H,H)	(H,M)	(H,H)
to	(H,M)	(M,M)	(M,M)
$\Delta\pi^R$	-262	-308	-570
	[2.50]	[0.95]	[2.44]
$\Delta\pi^M$	1,657	1,338	2,995
	[17.64]	[4.25]	[21.09]
$\Delta\pi^{M+R}$	1,395	1,030	2,425
	[15.63]	[4.25]	[19.23]
$\Delta\pi^H$	-868	-1,299	-2,167
	[4.79]	[8.90]	[13.59]
		Rebates	
Feasible	262 -1657	308-1338	570-2995
Observed	1,882	214	2,096
ΔPS	501	-272	229
	[17.39]	[10.75]	[27.53]
ΔCS	261	-110	150
	[11.44]	[6.39]	[16.55]
ΔSS	762	-383	379
	[28.56]	[16.94]	[43.90]

Notes: The top pane reports revenues under three assortments using an e^R effort policy for each one. The second pane reports changes in variable profit from moving from one assortment to another, as indicated. Rebate ranges in the third pane reflect the IR and IC constraints of the retailer and Mars. Standard errors are computed according to the procedure in Appendix A.6. The reported 'Observed' rebate uses the observed discount λ in the calculation of the rebate payment. ΔCS assumes a demand elasticity of $\epsilon = -2$. ΔPS includes revenues of Nestle, Mars, and Hershey.

Table 10: Critical Thresholds and Foreclosure at Observed λ

$\overline{\pi}_M^{MIN}$	$\overline{\pi}_M^{MAX}$	Assortment	Effort
0	11,763	(H,M)	$e^R(H,M)$
11,763	11,912	(H,M)	$e(\overline{\pi}_M(H,M))$
11,912	13,101	(M,M)	$e^R(M,M)$
13,101	13,319	(M,M)	$e(\overline{\pi}_M(M,M))$
13,320	∞	(H,H)	$e^{NR}(H,H)$

Notes: Calculations report the retailer's optimal assortment and effort policy at the observed λ for different values of the threshold.

Table 11: Optimal Effort Policies: Restock after how many customers?

	(H,H)	(H,M)	(M,M)
e^{NR}	263	267	264
e^R	257	261	259
e^{VI}	237	244	243
e^{IND}	241	247	244
$e^{SOC}(\epsilon = -4)$	233	238	235
$e^{SOC}(\epsilon = -2)$	227	232	229
$e^{SOC}(\epsilon = -1)$	220	224	222

Notes: Socially-optimal effort levels reported for different calibrated median own price elasticities of demand. For further details, see Appendix A.4. The width of the 95% CI is at most one unit.

Table 12: Potential Gains from Effort

	Vertically Integrated			Socially Optimal		
	(H,H)	(H,M)	(M,M)	(H,H)	(H,M)	(M,M)
$\%\Delta(e^{NR}, e^{Opt})$	9.89	8.61	7.95	13.69	13.11	13.26
$\%\Delta(e^{R}, e^{Opt})$	7.78	6.51	6.18	11.67	11.11	11.58
$\Delta\pi^R$	-83	-63	-55	-163	-152	-157
	[2.75]	[2.51]	[2.30]	[4.23]	[3.77]	[3.87]
$\Delta\pi^M$	195	152	128	251	211	190
	[5.83]	[5.10]	[4.92]	[6.61]	[5.62]	[5.70]
ΔPS	76	65	63	39	24	17
	[3.09]	[2.65]	[3.04]	[3.32]	[3.30]	[3.64]
$\Delta CS(\epsilon = -2)$	228	210	192	289	290	284
	[5.74]	[5.68]	[5.88]	[5.93]	[5.65]	[6.08]
ΔSS	304	275	255	329	313	301
	[8.51]	[8.02]	[8.63]	[8.45]	[7.78]	[8.54]

Notes: Percentage change in policy is calculated as increase required from baseline policy e^{NR} to vertically integrated or socially optimal policy. Social optimum assumes α corresponding to a median own price elasticity of demand of $\epsilon = -2$. For robustness, see Appendix A.4.

Table 13: Net Effect of Efficiency and Foreclosure

Base:	(H,H) and e^{NR}			(H,M) and e^{NR}		
to (M,M) with effort:	e^R	e^{VI}	e^{SOC}	e^R	e^{VI}	e^{SOC}
$\Delta\pi^R$	-575	-626	-728	-312	-364	-466
	[2.39]	[2.75]	[4.21]	[0.93]	[2.31]	[3.79]
$\Delta\pi^M$	3,045	3,140	3,201	1,382	1,476	1,538
	[21.59]	[21.74]	[22.07]	[4.88]	[6.05]	[6.72]
ΔPS	267	302	255	-239	-203	-250
	[27.84]	[27.58]	[27.03]	[10.71]	[10.30]	[9.81]
ΔCS $(\epsilon = -2)$	211	352	444	-49	92	185
	[17.14]	[18.07]	[19.32]	[7.03]	[7.92]	[8.95]
ΔSS	477	654	700	-287	-111	-65
	[44.86]	[45.38]	[46.03]	[17.43]	[17.63]	[18.20]

Notes: Consumer Surplus calibrates α to median own price elasticity of $\epsilon = -2$. Calibration only affects the scale of consumer surplus calculations, not the ranking of various options. For more details see Appendix A.4. Only one of our 1000 bootstrap iterations (ΔSS for the $e^{SOC}(H,M)$ case) yields a different sign than those reported in the table.

Table 14: Potential Upstream Deviations

Base: to (M,M) with effort:	(H,H) and e^{NR} e^R	e^{VI}	e^{SOC}	(H,M) and e^{NR} e^R	e^{VI}	e^{SOC}
$\Delta\pi^R$	-575	-626	-728	-312	-364	-466
$\Delta\pi^M$	3,045	3,140	3,201	1,382	1,476	1,538
$\Delta\pi^H$	-2,173	-2,173	-2,173	-1,302	-1,302	-1,302
$\lambda\pi^M$	2,096	2,111	2,121	2,096	2,111	2,120
w_h to avoid foreclosure	12.83	13.54	15.35	-15.83	-14.61	-11.59
	[0.23]	[0.24]	[0.23]	[0.52]	[0.53]	[0.51]
Reduction in λ ($w_h = \$0.15$)	5.27%	3.53%	-0.84%	44.79%	42.72%	38.18%
	[0.52]	[0.55]	[0.55]	[0.37]	[0.41]	[0.43]

Notes: Hershey's wholesale price to avoid foreclosure (w_h) is in cents per unit. The reduction in the discount assumes $w_h = 0.15$.

Table 15: Linear Pricing vs. AUD (Assortment is (M,M))

	e^R	e^{VI}	Linear Pricing
$\overline{\pi}^M$	$\in$ [11912,13101]	=13,195	=0
e	259	243	257
$\pi^R + \lambda\pi^M$	38,182	38,146	39,103
$(1-\lambda)\pi^M$	11,005	11,084	10,094
PS	50,441	50,476	50,450
CS ($\epsilon = -2$)	24,812	24,953	24,832

Notes: The optimal wholesale price under linear pricing is estimated to be 41.36 cents per unit. Hershey is excluded in the (M,M) assortment for all three arrangements, and earns zero profit. The changes in producer surplus include small changes in Nestle's profits due to the effect of changes in the retailer's choice of restocking policy on the sales of Raisinets.

Table 16: Comparison under Alternate Ownership Structures

	No Merger	M-H Merger	M-N Merger	H-N Merger
AUD Assortment	$e^{VI}(M,M)$	$e^{VI}(H,M)$	$e^{VI}(M,M)$	$e^{VI}(M,M)$
Alternative	$e^{NR}(H,H)$	$e^{NR}(N,N)$	$e^{NR}(H,H)$	$e^{NR}(H,H)$
$\Delta\pi^R$	-626	-253	-616	-626
	[4.11]	[6.15]	[3.80]	[4.11]
$\Delta\pi^M$	3,140	2,962	3,091	3,140
	[22.10]	[15.25]	[20.90]	[22.10]
$\Delta\pi^{Rival}$	-2,173	-1,458	-2,173	-2,212
	[13.60]	[1.47]	[13.60]	[12.75]
$\lambda\pi^M$	2,111	2,105	2,309	2,111
	[4.71]	[3.44]	[4.45]	[4.71]
ΔPS	302	1251	302	302
	[27.48]	[12.35]	[27.61]	[27.48]
ΔCS ($\epsilon = -2$)	352	769	337	352
	[18.98]	[9.47]	[18.95]	[18.98]
Price to Avoid Foreclosure	13.53	-11.90	9.44	14.04
	[0.22]	[0.15]	[0.25]	[0.21]
% Reduction in Rebate ($c = 0.15$)	3.55	42.33	12.24	2.34
	[0.51]	[0.26]	[0.46]	[0.49]

Notes: Table compares the welfare impacts of an exclusive Mars stocking policy under alternative ownership structures. This assumes threshold is set at the vertically-integrated effort level.

Appendix

A.1: Proof of Theorems

Proof of Theorem 1:
Note: We can relate our (linear) delta operators to one another via:

$$\Delta\pi^* = \Delta_M\pi^* + \Delta_H\pi^*$$

(A3) provides that $\pi^I(M,M) > \pi^I(H,H)$. (B3) provides $\pi^I(H,M) > \pi^I(H,H)$ and (C4) provides that $\pi^I(H,M) > \pi^I(M,M)$. Thus $\pi^I(H,M) > \pi^I(M,M) > \pi^I(H,H)$.

Absent transfers, if R selects the assortment then $\pi^R(H,H) > \pi^R(H,M) > \pi^R(M,M)$ implies that the equilibrium assortment will be (H,H). If we temporarily ignore (H,M) then (A1)-(A3) say that in a choice between (M,M) and (H,H) it is possible to design a transfer T which leads to assortment $(H,H) \rightarrow (M,M)$ in equilibrium. Likewise, if we temporarily ignore (M,M), then under (B1)-(B3) it is possible to design a transfer that leads to assortment $(H,H) \rightarrow (H,M)$ in equilibrium. This narrows it down to two possible equilibria: $\{(M,M),(H,M)\}$.

If M gets to choose the contract and the transfer T then for (M,M) to be the equilibrium outcome, it remains to show that:

$$\pi^M(M,M) - T \quad \geq \quad \pi^M(H,M) - T_h$$

The dominant manufacturer M should choose the smallest such T in each case so that (A1) or (B1) binds.

$$\begin{aligned}
\pi^M(M,M) - (-\Delta\pi^R) &\geq \pi^M(H,M) - (-\Delta_H\pi^R) \\
\pi^M(M,M) + \Delta\pi^R &\geq \pi^M(H,M) + \Delta_H\pi^R \\
\underbrace{\pi^M(M,M) - \pi^M(H,M)}_{\Delta_M\pi^M} + \underbrace{\Delta\pi^R - \Delta_H\pi^R}_{\Delta_M\pi^R} &\geq 0
\end{aligned}$$

This gives us the sensible condition that (M,M) is preferred by M to (H,M) when such a change would increase the bilateral surplus between M and R, which is guaranteed by (C4) $\square$.
Proof of Theorem 2:
Using (C3) instead of (C4) ensures that $\pi^I(H,M) < \pi^I(M,M)$. Thus $\pi^I(M,M) > \pi^I(H,M) > \pi^I(H,H)$. In the final line, (C3) guarantees that $\Delta_M\pi^M + \Delta_M\pi^R \geq -\Delta_M\pi^H \geq 0$. Thus (M,M) is the unique equilibrium and it maximizes overall industry profits $\square$.

A.2: Alternative Contracts

This section compares the AUD contract to other contractual forms; it is meant to be expositional and does not present new theoretical results.

Quantity Discount

A discount d, can be mapped into λ (a share of M's variable profit margin). However the discount no longer applies to all q_m, only those units in excess of the threshold, so that $\gamma(\overline{\pi^M}) = \max\left\{0, \frac{\pi^M - \overline{\pi^M}}{\pi^M}\right\}$. This implies $T \equiv \gamma(\overline{\pi^M}) \cdot \lambda \cdot \pi^M$, so that as the threshold increases, M is limited in how much surplus he can transfer to R, assuming that the post-discount wholesale price is non-negative. In the limiting case, the threshold binds exactly and M cannot offer R any surplus. This makes the discount, rather than the threshold, the primary tool for incentivizing effort. (Recall that for the AUD, $\overline{e} \geq e^R$ implies that M can directly set the retailer's effort). This means that high effort levels, $e > e^R$, will be more expensive to the dominant firm under the quantity discount than under the AUD. In fact, the vertically-integrated level of effort is only achievable through the 'sell out' discount, where $d = w_m - c_m$ such that M earns no profit on the marginal unit, and some $\overline{q}_m$ significantly less than the vertically-integrated quantity.

Quantity Forcing Contract

The quantity forcing (QF) contract is similar to a special case of the AUD contract. Specify a conventional AUD $(w_m, d, \overline{q}_m)$ as:

$$\begin{cases} (p_m - w_m + d) \cdot q_m & \text{if } q_m \geq \overline{q_m} \\ (p_m - w_m) \cdot q_m & \text{if } q_m < \overline{q_m} \end{cases}$$

One can increase the wholesale price w_m by one unit, and the generosity of the rebate (d) by one unit. Continuing with this procedure, the retailer profits when the threshold is met. For any $q_m \geq \overline{q}_m$, the retailer's profit remains unchanged, while his profit for any $q_m < \overline{q}_m$, tends to zero as $w_m \to p_m$. This has the effect of 'forcing' the retailer to accept a quantity at least as large as $\overline{q}_m$. By choosing the threshold, the QF contract can achieve the vertically-integrated level of effort, just like the AUD. For quantities $q_m > \overline{q}_m$, the AUD works like a QF contract plus a uniform wholesale price on 'extra' units.[78]. Without some outside constraint on d or w_m, and absent uncertainty about demand, the dominant firm has an incentive to increase d and w_m together to replicate the QF contract.

Two-Part Tariff

[78]For a more complete discussion of the connection between the AUD and the QF contract in the presence of a capacity constrained rival see Chao and Tan (2014)

One can also construct a two-part tariff (2PT), described by two terms: a share of M's revenue λ and a fixed transfer T from $R \to M$. The retailer chooses between the 2PT contract and the standard wholesale price contract.

$$\begin{cases} \pi^R(a,e) + \lambda \cdot \pi^M(a,e) - T & \text{if } 2PT \\ \pi^R(a,e) & \text{o.w.} \end{cases}$$

We define $\underline{\pi}^R = \max_{a,e} \pi^R(a,e)$ (the retailer's optimum under the standard wholesale price contract). For the retailer to choose the 2PT contract it must be that $\max_{a,e}\{\pi^R(a,e) + \lambda \cdot \pi^M(a,e) - T\} \geq \underline{\pi}^R$. An important case of the 2PT contract is the so-called 'sellout' contract where $\lambda = 1$. In this case, the retailer maximizes the joint surplus of $\pi^R + \pi^M$ and achieves the vertically-integrated assortment and stocking level. Just like in the AUD, this may lead to foreclosure of the rival H, even when that foreclosure is not optimal from an industry perspective. The dominant firm can choose T so that $\max_{a,e}\{\pi^R(a,e) + \pi^M(a,e)\} - T = \underline{\pi}_M$ and "fully extract" the surplus from R. Likewise, the dominant firm can choose $T = (1 - \lambda_{AUD}) \cdot \overline{\pi}^M$ (the dominant firm's profits under the AUD) so long as the retailer is willing to choose the 2PT contract.

This indicates that it is also possible for a 2PT contract to implement the assortment and effort level that maximizes the bilateral profit between $M + R$, even if that assortment does not maximize overall industry profits. An important question is: how do the AUD and the 2PT differ? One possibility is that the AUD can be used to implement an effort level in excess of the vertically-integrated optimal effort, e^{VI}, which results in higher profits for M at the expense of the retailer. A major challenge of devising a 2PT in practice is arriving at the fixed fee T, especially when there are multiple retail firms of different sizes, and the 2PT contract (or menu of contracts) is required to be non-discriminatory.[79] It may be easier in practice to tailor sales thresholds to the size of individual retailers (as opposed to setting individual fixed-fee transfer payments).[80]

A.3: Effort Derivation

Consider the effort choice of the retailer faced with an AUD contract from (2):

$$\begin{cases} \pi^R(e) - c(e) + \lambda \cdot \pi^M(e) & \text{if } \pi^M(e) \geq \overline{\pi^M} \\ \pi^R(e) - c(e) & \text{if } \pi^M(e) < \overline{\pi^M} \end{cases}$$

[79] Kolay, Shaffer, and Ordover (2004) shows that a menu of AUD contracts may be a more effective tool in price discriminating across retailers than a menu of 2PTs. In the absence of uncertainty, an individually-tailored 2PT enables full extraction by M, but is a likely violation of the Robinson-Patman Act.

[80] Another possibility as shown by O'Brien (2013) is that the AUD contract can enhance efficiency under the double moral-hazard problem (when the upstream firm also needs to provide costly effort such as advertising).

In the case where the rebate is paid, we can express the retailer's problem as:

$$e_1 = \arg\max_e \pi^R(e) - c(e) + \lambda\pi^M(e) \quad \text{s.t.} \quad \pi^M(e) \geq \overline{\pi^M}$$

The solution to the constrained problem is given by:

$$e_1 = \max\{e^R, \bar{e}\} \quad \text{where} \quad \bar{e} \text{ solves} \quad \pi^M(\bar{e}) = \overline{\pi^M}$$

If the rebate is not paid then:

$$e_0 = e^{NR} = \arg\max_e \pi^R(e) - c(e)$$

The retailer's IC constraint:

$$\pi^R(e_1) - c(e_1) + \lambda\pi^M(e_1) \geq \pi^R(e_0) - c(e_0) \tag{IC}$$

and the dominant firm M's IR constraint:

$$(1-\lambda)\pi^M(e_1) \geq \pi^M(e_0) \tag{IRM}$$

When we consider the sum of (IC) and (IRM) it is clear that a rebate which induces effort level e_1 must increase bilateral surplus relative to e_0:

$$\pi^R(e_1) - c(e_1) + \pi^M(e_1) \geq \pi^R(e_0) - c(e_0) + \pi^M(e_0)$$

This provides an upper bound on the effort that can be induced by the rebate contract.

A.4: Computing Treatment Effects

One goal of the exogenous product removals is to determine how product-level sales respond to changes in availability. Let q_{jt} denote the sales of product j in machine-week t, superscript 1 denote sales when a focal product(s) is removed, and superscript 0 denote sales when a focal product(s) is available. Let the set of available products be A, and let F be the set of products we remove. Thus, $Q_t^1 = \sum_{j \in A \setminus F} q_{jt}^1$ and $Q_s^0 = \sum_{j \in A} q_{js}^0$ are the overall sales during treatment week t, and control week s respectively, and $q_{fs}^0 = \sum_{j \in F} q_{js}^0$ is the sales of the removed products during control week s. Our goal is to compute $\Delta q_{jt} = q_{jt}^1 - E[q_{jt}^0]$, the treatment effect of removing products(s) F on the sales of product j.

There are two challenges in implementing the removals and interpreting the data generated by them. The first challenge is that there is a large amount of variation in overall sales at the weekly level, independent of our exogenous removals. 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. In our particular setting, many of the product removals were done during the summer of 2007, which was a high-point in demand at these sites, most likely due to macroeconomic conditions. In this case, using a simple measure like previous weeks' sales, or overall average sales for $E[q_{jt}^0]$ could result in unreasonable treatment effects, such as sales increasing due to product removals, or sales decreasing by more than the sales of the focal products.

In order to deal with this challenge, we impose two simple restrictions based on consumer theory. Our first restriction is that our experimental product removals should not increase overall demand, so that $Q_t^0 - Q_s^1 \geq 0$ for treatment week t and control week s. Our second restriction is that the product removal(s) should not reduce overall demand by more than the sales of the products we removed, or $Q_t^0 - Q_s^1 \leq q_{fs}^0$. This means we choose control weeks s that correspond to treatment week t as follows:

$$\{s : s \neq t, Q_t^0 - Q_s^1 \in [0, q_{fs}^0]\}. \tag{12}$$

While this has the nice property that it imposes the restriction on our selection of control weeks that all products are weak substitutes, it has the disadvantage that it introduces the potential for selection bias. The bias results from the fact that weeks with unusually high sales of the focal product q_{fs}^0 are more likely to be included in our control. This bias would likely overstate the costs of the product removal, which would be problematic for our study.

We propose a slight modification of (12) which removes the bias. That is, we replace q_{fs}^0 with $\widehat{q_{fs}^0} = E[q_{fs}^0 | Q_s^0]$. An easy way to obtain the expectation is to run an OLS regression of q_{fs}^0 on Q_s^0, at the machine level, and use the predicted value. This has the nice property that the error is orthogonal to Q_s^0, which ensures that our choice of weeks is unbiased.

The second challenge is that, although the experimental design is relatively clean, the product mix presented in a machine is not necessarily fixed across machines, or within a machine over long periods of time, because we rely on observational data for the control weeks. For example, manufacturers may change their product lines, or Mark Vend may change its stocking decisions over time. Thus, while our field experiment intends to isolate the treatment effect of removing Snickers, we might instead compute the treatment effect of removing Snickers jointly with Mark Vend changing pretzel suppliers.

To mitigate this issue, we restrict our set of potential control weeks to those at the same machine with similar product availability within the category of our experiment. In practice, two of our three treatments took place during weeks where 3 Musketeers and Reese's Peanut Butter Cups were unavailable, so we restrict our set of potential control weeks for those experiments to weeks where those products were also unavailable. We denote this condition as $A_s \approx A_t$.

We use our definition of control weeks s to compute the expected control sales that correspond to treatment week t as:

$$S_t = \{s : s \neq t, A_t \approx A_s, Q_t^0 - Q_s^1 \in [0, \hat{b_0} + \hat{b_1} Q_s^0]\}. \quad (13)$$

And for each treatment week t we can compute the treatment effect as

$$\Delta q_{jt} = q_{jt}^1 - \frac{1}{\#S_t} \sum_{s \in S_t} q_{js}^0. \quad (14)$$

While this approach has the advantage that it generates substitution patterns consistent with consumer theory, it may be the case that for some treatment weeks t the set of possible control weeks $S_t = \{\emptyset\}$. Under this definition of the control, some treatment weeks constitute 'outliers' and are excluded from the analysis. Of the 1470 machine-experiment-week combinations, 991 of them have at least one corresponding control week, and at the machine-experiment level, 528 out of 634 have at least one corresponding control. Each included treatment week has an average of 24 corresponding control weeks, though this can vary considerably from treatment week to treatment week.[81]

Once we have constructed our restricted set of treatment weeks and the set of control weeks that corresponds to each, inference is fairly straightforward. We use (14) to construct a set of pseudo-observations for the difference, and employ a paired t-test.

A.5: Empirical Distribution of Arrival Rate

There is heterogeneity in both the arrival rate of consumers to machines, as well as the service level of different machines in the data. In order to construct our 'representative machine,' we divide our sample of 66 experimental machines into four groups using a k-means clustering algorithm based on the arrival rate, and the amount of revenue collected at a service visit. We report summary statistics from the four groups in table 17. Our counterfactual analyses are based on cluster D, which is the largest cluster, containing 28 of the 66 machines in our sample. Machines in clusters A and C are smaller in size, while the seven machines in cluster B represent the very highest volume machines in the sample. We focus on cluster D because it is a large cluster of 'higher than average volume' machines, which we think is the most important determinant of the retailer's re-stocking decision. The distribution of daily sales for the machines in cluster D determine the transition rule for our re-stocking model. Figure 5 plots a histogram of daily sales for the machines in cluster D.

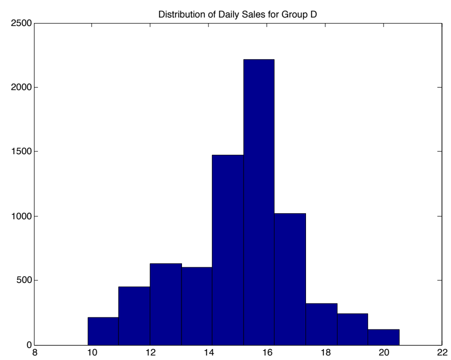
[81] Weeks in which the other five treatments were run (for the salty-snack and cookie categories) are excluded from the set of potential control weeks.

Table 17: Summary of Sales and Revenues for Four Clusters of Machines

	Group Size	Vends/Visit		Revenue/Visit		Avg Sales/Day	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
A	4	39.0	26.1	28.3	18.7	5.8	1.4
B	7	88.9	39.5	70.6	33.4	24.9	3.0
C	27	56.9	31.5	41.5	23.2	9.2	1.4
D	28	71.6	33.8	54.3	26.8	15.1	2.0

Notes: The 66 machines in our analyses are divided into four groups of machines based on the arrival rate and the amount of revenue collected at a service visit, using a k-means clustering algorithm. Our counterfactual analyses are based on cluster D.

Figure 5: Histogram of Daily Sales for Machines in Group D



Notes: The 28 machines in group D form the basis for our counterfactual exercises. Means and standard deviations for all machine groups are reported in table 17.

A.6: Details for Computation of Standard Errors

Here we describe the procedure used to calculate the standard errors produced throughout the paper.

1. Draw $\hat{\theta}^b \sim N\left(\hat{\theta}^{MLE}, \sqrt{diag(V(\hat{\theta}^{MLE})}\right)$.

2. Use the same (calibrated) arrival process for $f(x'|x)$, same calibrated discount factor β and same calibrated restocking cost $FC = 10$.

3. Normalize the state space using the outside good share of a hypothetical assortment containing all chocolate products a^*: $\frac{1}{1-s_0(a^*,\hat{\theta}^b)}$.

4. For each assortment a, simulate consumer arrivals to obtain $u^R(x,a|\hat{\theta}^b)$,$u^M(x,a|\hat{\theta}^b)$, etc. (Repeat 100,000 times).

5. Fit a Chebyshev Polynomial (order 10) to the average of each computed sequence of profits: $u^R(x,a|\hat{\theta}^b)$,$u^M(x,a|\hat{\theta}^b)$, etc.

6. For every possible value of e use (11) to compute:
$\pi^*(a,e|\hat{\theta}^b) = \Gamma(\tilde{P}(e)) \cdot (I - \beta\tilde{P}(e))^{-1} \cdot \hat{u}^*(x,a|\hat{\theta}^b))$.

7. Use $\pi^*(a,e|\hat{\theta}^b)$ to calculate the optimal policies for different groups of agents (e^{NR},e^R,e^{VI},e^{SOC}) for every a.

8. Compute all of the profit differences $\Delta\pi^R, \Delta\pi^M, \Delta\pi^H$ for Tables 9-15.

9. Repeat 1000 times and report the standard deviations.

In this procedure there are two sources of variation. The first is the variation introduced by the uncertainty in the MLE estimates of the demand parameters (as reported in Table 7). The second is the simulation variance introduced from our simulation procedure, because we use the average over 100,000 chains this is designed to be at most $\pm\$2$.

A.7: Consumer Surplus and Welfare Calculations

Our calculation of the expected consumer surplus of a particular assortment and effort policy (a,e) parallels our calculation of retailer profits. We simulate consumer arrivals over many chains, and compute the set of available products as a function of the initial assortment a and the number of consumers to arrive since the previous restocking visit x which we write $a(x)$. For each assortment $a(x)$ that a consumer faces, we can compute the logit inclusive value and average over our simulations, to obtain an estimate at each x:

$$CS^*(a,x|\theta) = \frac{1}{NS}\sum_{s=1}^{NS} \log\left(\sum_{j\in a(x^s)} \exp[\delta_j + \mu_{ij}(\theta)]\right)$$

The exogenous arrival rate, $f(x'|x)$, denotes the expected daily number of consumer arrivals (from x cumulative likely consumers today to x' cumulative likely consumers tomorrow). Using this arrival rate and a policy $x^*(e)$, we obtain the post-decision transition rule $\tilde{P}(x^*(e))$ and evaluate the ergodic distribution of consumer surplus under policy e:

$$CS^*(a,e) = (I - \beta\tilde{P}(x^*(e)))^{-1} CS^*(a,x|\theta)$$

The remaining challenge is that $CS^*(a,e)$ relates to arbitrary units of consumer utility, rather than dollars. Recall our utility specification from (4), with $\theta = [\delta,\alpha,\sigma]$:

$$u_{ijt}(\theta) = \delta_j + \alpha p_{jt} + \xi_t + \sum_l \sigma_l \nu_{ilt} x_{jl} + \varepsilon_{ijt}$$

Without observable, within-product variation in price, $p_{jt} = p_j$, and α is not separately identified from the product fixed-effect δ_j. If α were identified, then we could simply write $CS(a,e) = \frac{1}{\alpha}CS^*(a,e)$. Instead, we can calibrate α given an own price elasticity:

$$\epsilon_{j,t} = \frac{p_{jt}}{s_{jt}} \cdot \frac{\partial s_{jt}}{\partial p_{jt}} = \frac{p_{jt}}{s_{jt}} \cdot \int \frac{\partial s_{ijt}}{\partial p_{jt}} f(\beta_i|\theta) d\,\beta_i = \alpha \cdot \underbrace{\frac{p_{jt}}{s_{jt}} \cdot \int (1 - s_{ij}(\delta,\beta_i)) \cdot s_{ij}(\delta,\beta_i) f(\beta_i|\theta) d\,\beta_i}_{\epsilon^*_{j,t}(\theta)}$$

The term $\epsilon^*_{j,t}$ does not depend directly on α once we have controlled for the fixed effect d_j. Thus, we can calibrate own-price elasticities. As is conventional in the literature, we work with the median own-price elasticity, $\overline{\epsilon}(\theta) = \text{median}_j(\epsilon^*_{j,t}(\theta))$, and recover α as $\alpha = |\frac{\epsilon}{\overline{\epsilon}(\theta)}|$. We then calculate α at three different values of the median own price elasticity: $\epsilon \in \{-1, -2, -4\}$.

As is well known, α has an alternative interpretation in the social planner's problem as the planner's weight on consumer surplus:

$$SS(a,e) = PS(a,e) + \frac{1}{\alpha} CS^*(a,e)$$

The social planner's problem is equivalent in the following cases: (1) the median own-price elasticity is $\epsilon = -2$; (2) the median own-price elasticity is $\epsilon = -4$ and the planner puts twice as much weight on consumer surplus; (3) the median own-price elasticity is $\epsilon = -1$ and the planner puts half as much weight on consumer surplus.

In the following table 18, we show robustness to assumptions about the median own price elasticity of demand (and the corresponding α parameter). As consumers become more inelastic, the social planner places more weight on consumer utility in calculating the socially-optimal stocking rule. The socially-optimal stocking policy is to restock after $(220 - 224)$ consumers (depending on the assortment). This represents an approximately 16% decrease in the number of consumers between restocking visits when compared to the no-rebate case. The potential gains to consumer surplus are large ($\Delta SS \geq \$600$), and total producer surplus falls. When consumer demand is more elastic (median own-price elasticity $\epsilon = -4$) the socially-optimal policy is in the range of $(233-238)$, representing a 10-12% reduction in the number of consumers between visits. The social planner places less weight on consumer surplus, and the gains from implementing the socially-optimal stocking policy are smaller ($\Delta SS \leq 200$).

The case that we report in the text of the paper assumes that median own-price elasticity is $\epsilon = -2$. We choose this because it is relatively inelastic when compared to own-price elasticities recovered from demand systems in the literature, and is thus meant to represent an 'upper bound' on potential efficiency effects from increased restocking.

Table 18: Socially Optimal Effort Policies (under various elasticities)

	$\epsilon=-1$			$\epsilon=-2$			$\epsilon=-4$		
e^{SOC}	220	224	222	227	232	229	233	238	235
$\%\Delta(e^{NR}, e^{SOC})$	16.35	16.10	15.91	13.69	13.11	13.26	11.41	10.86	10.98
$\%\Delta(e^{R}, e^{SOC})$	14.40	14.18	14.29	11.67	11.11	11.58	9.34	8.81	9.27
$\Delta\pi^R$	-238	-234	-230	-163	-152	-157	-112	-102	-106
$\Delta\pi^M$	285	242	213	251	211	190	219	183	166
ΔPS	-12	-35	-36	39	24	17	66	51	46
ΔCS	645	659	637	289	290	284	128	126	124
ΔSS	633	624	601	329	313	301	193	178	170

Notes: Reports effort policies that maximize social surplus, under different assumptions for median own-price elasticity when calculating consumer surplus.

A.8: Retailer Optimizes Retailer/Consumer Joint Surplus

As a robustness test, we allow the retailer to jointly optimize the joint surplus of the retailer and the consumer. This may be an important consideration if providing good service to the consumer is an important aspect of how our retail operator competes with other vending operators for contracts with retail locations. It may also help explain why our retailer provides an extremely high frequency of service visits (beyond what we can justify with an optimal stocking model).

Table 19 reports the optimal effort policies of a joint Retailer-Consumer entity. The main distinction is that the retailer exerts far more effort when maximizing his own profit when compared to when he did not take consumer surplus into account. This has the effect of substantially reducing the gap between the social optimum e^{SOC} and the retailer optimum (absent the rebate) e^{NR} to 9 consumers or less in the $\epsilon=-2$ case. The gap is smaller as consumers become more inelastic at most 5 consumers for the $\epsilon=-1$ case, and larger as consumers become more elastic as many as 12 in the $\epsilon=-4$ case. Once retailers take into account consumer surplus, there is no longer a distinction between the industry optimal policy and the socially optimal policy. Also, the gap between the vertically integrated policy and the socially optimal (or industry optimal) policy depends only on the profits of the competing firms, and is generally 3 customers or fewer.

We report the potential gains from socially optimal effort levels for the joint Retailer-Consumer in Table 20. The potential gains are much smaller than they are in the case where the retailer does not take consumer surplus into account. For all elasticities, the potential change in the restocking frequency is now less than 5%. Likewise, the maximum change in social surplus is less than \$75 for all elasticities and assortments. Once the retailer internalizes the effect of effort on consumers, there is little to be gained from internalizing the same effort effect on the upstream manufacturer. The retailer-consumer pair sets exerts more effort than the vertically integrated retailer-Mars pair

Table 19: Effort Decisions of Joint Retailer-Consumer

	$\epsilon = -1$			$\epsilon = -2$			$\epsilon = -4$		
e^{NR}	225	228	226	236	239	237	245	249	247
e^{R}	224	227	225	234	237	235	242	246	244
e^{VI}	219	223	221	225	230	229	230	236	234
e^{IND}	220	224	222	227	232	229	233	238	235
e^{SOC}	220	224	222	227	232	229	233	238	235

Notes: Reports effort policies that maximize the combined retailer-consumer surplus, under different assumptions for median own-price elasticity when calculating consumer surplus.

Table 20: Socially Optimal Effort Policies (Joint Retailer-Consumer)

	$\epsilon = -1$			$\epsilon = -2$			$\epsilon = -4$		
$\%\Delta(e^{NR}, e^{OPT})$	2.22	1.75	1.77	3.81	2.93	3.38	4.90	4.42	4.86
$\%\Delta(e^{R}, e^{OPT})$	1.79	1.32	1.33	2.99	2.11	2.55	3.72	3.25	3.69
$\Delta\pi^{R}$	-10	-6	-7	-19	-13	-16	-29	-23	-26
$\Delta\pi^{M}$	23	14	13	50	33	33	77	60	59
ΔPS	7	5	4	19	13	13	31	26	27
ΔCS	46	37	38	54	44	49	43	41	44
ΔSS	53	42	42	73	57	62	75	67	71

Notes: Reports potential gains realized when effort is chosen to maximize combined retailer-consumer surplus, under different assumptions for median own-price elasticity when calculating consumer surplus.

in our base scenario.

In Table 21, we calculate the optimal assortment decision of a joint Retailer-Consumer pair. We find that the assortment choice depends on how much weight the retailer places on consumer surplus, or how elastic consumers are. Assuming the retailer places full weight on consumer surplus, at a median own price elasticity of $\epsilon = -2$ the retailer is more or less indifferent between the (H,M) assortment and the (H,H) assortment. As consumers become more elastic, the retailer-consumer pair prefers (H,H), and as they become less elastic the retailer-consumer pair prefers the consumer-optimal assortment (H,M).

We combine foreclosure and efficiency effects where we treat the retailer-consumer as a jointly maximizing pair in Table 22. When consumers are sufficiently inelastic, and the retailer accounts for consumer utility when choosing the assortment, he selects (H,M). In this world, any rebate which induces a switch to (M,M) decreased both producer and consumer surplus. As consumers become more elastic (or the retailer places less weight on consumer surplus) the retailer chooses (H,H) absent the rebate, and the results qualitatively match the results in the main text: the rebate can increase both producer and consumer surplus relative (H,H) but full foreclosure is inefficient

Table 21: Joint Retailer-Consumer Optimal Assortment (under various elasticities)

	(H,M)	(H,H)	(M,M)
Median Elasticity $\epsilon = -1$			
π^R	36,176	36,507	35,904
CS	50,372	49,771	50,122
$\pi^R + CS$	**86,548**	86,278	86,026
Median Elasticity $\epsilon = -2$			
π^R	36,281	36,591	35,998
CS	25,126	24,815	24,997
$\pi^R + CS$	**61,407**	**61,406**	60,995
Median Elasticity $\epsilon = -4$			
π^R	36,340	36,637	36,053
CS	12,532	12,368	12,461
$\pi^R + CS$	48,872	**49,005**	48,514

Notes: Reports gains to retailer and consumers under different assortments, and under different assumptions for median own-price elasticity when calculating consumer surplus. The optimal assortment under each of the three assumptions for median elasticity is shown in boldface type.

in that it fails to implement the optimal (H,M) assortment, and efficiency gains from additional stocking (smaller when considering retailer-consumers jointly) are not sufficient to compensate for foreclosure.

Though it is likely in practice that our retailer at least partially considers consumer surplus when choosing his effort level, our base scenario ignores this possibility. Incorporating consumer surplus in the retailer's effort decision drastically reduces potential efficiency effects of the rebate contract. Ultimately, we interested in whether some efficiency effect might outweigh potential foreclosure effects, and we design our baseline estimates to be an "upper bound" on such effects.

A.9: Full $\pi(a,e)$ Tables

We compute $\pi(a,e)$ for every agent and 15 assortments. We report only the most relevant assortments and effort levels below. Note that $\pi(a,e)$ denotes the present discounted value of profits from a single representative machine. Annualized, enterprise-level profits are approximately 20-50 times larger.

Table 22: Joint Retailer-Consumer Net Foreclosure/Efficiency Effect

	$\epsilon = -1$	$\epsilon = -2$	$\epsilon = -2$	$\epsilon = -4$	$\epsilon = -4$
From	$e^{NR}(H,M)$	$e^{NR}(H,M)$	$e^{NR}(H,H)$	$e^{NR}(H,M)$	$e^{NR}(H,H)$
To	$e^{VI}(M,M)$	$e^{VI}(M,M)$	$e^{VI}(M,M)$	$e^{VI}(M,M)$	$e^{VI}(M,M)$
$\Delta\pi^R$	-329	-348	-658	-357	-654
$\Delta\pi^M$	1326	1345	3019	1368	3064
$\Delta\pi^H$	-1280	-1285	-2151	-1290	-2160
ΔPS	-286	-293	177	-287	215
ΔCS	-203	-81	230	-27	137
ΔSS	-490	-374	407	-313	351

Notes: Reports changes under different assumptions for median own-price elasticity when calculating consumer surplus.

Table 23: Profits under Alternate Product Assortments and Stocking Policies

Policy	π^R	$\lambda\pi^M$	π^M	π^H	π^N	$\pi^R + \pi^M$	PS	CS
(H,M) Assortment: Reeses Peanut Butter Cup and Three Musketeers								
e^{NR} (267)	36,399	1,875	11,719	1,302	1,260	48,117	**50,679**	**24,861**
	[24.1]	[4.4]	[27.4]	[8.9]	[3.7]	[18.3]	[18.1]	[138.9]
e^{R} (261)	36,394	1,882	11,763	1,299	1,257	48,157	**50,713**	**24,923**
	[24.1]	[4.3]	[27.0]	[8.9]	[3.7]	[18.8]	[18.6]	[139.9]
e^{VI} (244)	36,335	1,899	11,871	1,290	1,249	48,206	**50,744**	**25,071**
	[22.7]	[4.1]	[25.9]	[8.8]	[3.7]	[19.5]	[19.3]	[142.6]
(H,H) Assortment: Reeses Peanut Butter Cup and Payday								
e^{NR} **(263)**	**36,661**	1,609	**10,055**	**2,173**	**1,285**	**46,716**	50,174	24,601
	[23.1]	[2.4]	[14.8]	[13.6]	[3.5]	[14.3]	[24.6]	[131.3]
e^{R} (257)	36,656	1,617	10,106	2,167	1,282	46,762	50,211	24,662
	[23.0]	[2.2]	[14.0]	[13.6]	[3.5]	[14.6]	[24.7]	[132.6]
e^{VI} (237)	36,578	1,640	10,251	2,149	1,272	46,829	50,250	24,830
	[21.8]	[1.9]	[11.7]	[13.5]	[3.5]	[15.3]	[25.1]	[135.5]
(M,M) Assortment: Three Musketeers and Milkyway								
e^{NR} (264)	36,090	2,091	13,067	0	1,256	49,156	50,412	24,761
	[24.2]	[4.9]	[30.9]	[0.0]	[3.8]	[21.6]	[18.1]	[141.4]
e^{R} **(259)**	**36,086**	**2,096**	**13,101**	**0**	**1,254**	**49,187**	50,441	24,812
	[24.1]	[4.9]	[30.4]	[0.0]	[3.8]	[22.1]	[18.6]	[142.6]
e^{VI} **(243)**	**36,035**	**2,111**	**13,195**	**0**	**1,246**	**49,230**	50,476	24,953
	[22.8]	[4.7]	[29.3]	[0.0]	[3.8]	[22.6]	[19.0]	[145.1]

Notes: Profit numbers represent the long-run expected profit from a 'representative' machine. Rebate payments are assumed to only be paid under an (M,M) assortment; rebate payments under other assortments are reported in light typeface, but are assumed to not be paid to the retailer. The retailer's optimal assortment under each effort policy is reported in boldface type. The socially-optimal assortment is (H,M); we denote this with boldface type for the PS and CS columns.

Table 24: Profits after Mars-Hershey Merger

Policy	π^R	$\lambda\pi^M$	$\pi^M+\pi^H$	π^N	$\pi^M+\pi^H+\pi^R$	PS	CS
(H,M) Assortment: Reeses Peanut Butter Cup and Three Musketeers							
e^{NR} (267)	36,399	2,083	13,021	1,260	49,419	**50,679**	**24,861**
e^{R} (262)	**36,395**	**2,089**	**13,055**	**1,257**	**49,451**	**50,708**	**24,913**
e^{VI} (245)	**36,340**	**2,105**	**13,155**	**1,249**	**49,496**	**50,745**	**25,064**
(N,N) Assortment: Butterfinger and Crunch							
e^{NR} (257)	**36,594**	1,631	**10,193**	**2,707**	**46,787**	49,494	24,295
e^{R} (251)	36,589	1,639	10,246	2,700	46,835	49,535	24,355
e^{VI} (232)	36,514	1,662	10,386	2,681	46,900	49,581	24,512

Notes: Profit numbers represent the long-run expected profit from a 'representative' machine. Rebate payments are assumed to only be paid under an (H,M) assortment; rebate payments in light typeface are assumed to not be paid to the retailer.

Table 25: Profits after Mars-Nestle Merger

Policy	π^R	$\lambda\pi^M$	$\pi^M+\pi^N$	π^H	$\pi^M+\pi^N+\pi^R$	PS	CS
Reeses Peanut Butter Cup (H), Three Musketeers (M)							
e^{NR} (267)	36,399	2,077	12,978	1,302	49,377	**50,679**	**24,861**
e^{R} (262)	36,395	2,082	13,013	1,299	49,409	**50,708**	**24,913**
e^{VI} (245)	36,340	2,098	13,114	1,290	49,455	**50,745**	**25,064**
Reeses Peanut Butter Cup (H), Payday (H)							
e^{NR} (263)	**36,661**	1,815	**11,341**	**2,173**	**48,001**	50,174	24,601
e^{R} (257)	36,656	1,822	11,388	2,167	48,045	50,211	24,662
e^{VI} (239)	36,591	1,842	11,511	2,151	48,102	50,253	24,815
Three Musketeers (M), Milkyway (M)							
e^{NR} (264)	36,090	2,292	14,323	0	50,412	50,412	24,761
e^{R} (259)	**36,086**	**2,297**	**14,354**	**0**	**50,441**	50,441	24,812
e^{VI} (244)	**36,040**	**2,310**	**14,436**	**0**	**50,476**	50,476	24,946

Notes: Profit numbers represent the long-run expected profit from a 'representative' machine. Rebate payments are assumed to only be paid under an (M,M) assortment; rebate payments in light typeface are assumed to not be paid to the retailer.

Table 26: Profits after Hershey-Nestle Merger

Policy	π^R	$\lambda\pi^M$	π^M	$\pi^H+\pi^N$	$\pi^M+\pi^H+\pi^R$	PS	CS
Reeses Peanut Butter Cup (H), Three Musketeers (M)							
e^{NR} (267)	36,399	1,875	11,719	2,562	48,117	**50,679**	**24,861**
e^{R} (261)	36,394	1,882	11,763	2,556	48,157	**50,713**	**24,923**
e^{VI} (244)	36,335	1,899	11,871	2,538	48,206	**50,744**	**25,071**
Reeses Peanut Butter Cup (H), Payday (H)							
e^{NR} (263)	**36,661**	1,609	**10,055**	**3,458**	**46,716**	50,174	24,601
e^{R} (257)	36,656	1,617	10,106	3,449	46,762	50,211	24,662
e^{VI} (237)	36,578	1,640	10,251	3,421	46,829	50,250	24,830
Three Musketeers (M), Milkyway (M)							
e^{NR} (264)	36,090	2,091	13,067	1,256	49,156	50,412	24,761
e^{R} (259)	**36,086**	**2,096**	**13,101**	**1,254**	**49,187**	50,441	24,812
e^{NR} (243)	**36,035**	**2,111**	**13,195**	**1,246**	**49,230**	50,476	24,953

Notes: Profit numbers represent the long-run expected profit from a 'representative' machine. Rebate payments are assumed to only be paid under an (M,M) assortment; rebate payments in light typeface are assumed to not be paid to the retailer.