Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee

Brian McManus*

May 2006

Abstract

Firms that practice second-degree price discrimination may distort product characteristics away from their efficient levels (*e.g.*, the small version of a product is "too small"). This paper offers the first empirical study of this product design issue. Using data from a specialty coffee market, I estimate a structural utility model that allows for consumer screening under vertical preference heterogeneity. Comparisons of cost data and the estimated benefits from changing product characteristics confirm some of the central predictions of nonlinear pricing theory. Product design distortions are relatively large for drinks that are not the most profitable but over which the firms hold market power. As drink size increases, the estimated distortions decrease toward zero for the drinks with the largest price-cost margins. This result provides empirical support for the "no distortion at the top" prediction from theory.

*Olin School of Business, Washington University, St. Louis, MO 63130. Email: mcmanus@wustl.edu. I would like to thank Simon Anderson, Steve Stern, and Maxim Engers for many helpful comments and suggestions. Editor Rob Porter and anonymous referees also offered valuable advice, as did Andrew Cohen, Peter Davis, Gautam Gowrisankaran, Bart Hamilton, Phillip Leslie, Steve Levitt, and Raphael Thomadsen. Data and information on the University of Virginia (UVa) coffee market were generously provided by Diane McClellan and Susan Presto at Aramark/Greenberry's, John Salidis of Espresso Corner, Joseph and William Trager of Higher Grounds, and Robert Harris at Starbucks. Jeffrey Tilman at UVa's Facilities Management Department and Michael Furlough of the UVa Geospatial and Statistical Data Center provided information and analysis that were crucial in constructing consumer population data.

1 Introduction

Nonlinear price schedules are characterized by marginal prices that vary with product size or quality. The most common nonlinear prices are quantity discounts and multi-part tariffs. Nonlinear pricing appears in competitive markets when marginal or average costs change with product size, but firms with market power also may use nonlinear prices for second-degree price discrimination. When consumers hold private information about their tastes, a monopolist or oligopolist can use nonlinear pricing as a screening mechanism to induce different types of consumers to buy different products. The pricing strategy is generally coupled with product design decisions that determine how much of a product a consumer will receive. Screening incentives may lead a firm to make a small version of its product "too small" in order to collect more profit from consumers who purchase the larger version.

The modern theoretical second-degree price discrimination literature was initiated by Mussa and Rosen (1978) and Maskin and Riley (1984). (The abbreviation "MR" refers to the contributions of both pairs of authors.) MR consider a monopolist's design of a price-quantity[1] menu when consumers hold private information over a scalar taste parameter. The equilibrium product menu offers the highest demand consumer an efficiently sized product, and all other buyers self-select into inefficiently small purchasing options. The prices that induce this allocation often include quantity discounts that are unrelated to costs. Competition has been introduced to the MR setting in a variety of ways.[2] For my purposes, the most relevant models of oligopoly nonlinear pricing include a consumer population with private information over vertical and horizontal characteristics. In this case, Rochet and Stole (2002) and Armstrong and Vickers (2001) show that with sufficient (but not absolute) competition the equilibrium nonlinear price schedule is "cost-plus-fee" and product allocations are efficient. When competition is weakened, product menus are typically between the MR allocation and the efficient one.

In this paper I present a novel empirical study of product design efficiency under nonlinear pricing. I estimate the patterns of allocative distortions that follow from oligopolistic firms offering product menus with quantity discounting. I use the estimates to evaluate whether (and which) distortion patterns predicted in the theoretical literature are realized empirically. I consider the

[1] Product size and quality are generally interchangeable in the literature.

[2] The MR model is difficult to extend to settings with multidimensional consumer or competition. See Wilson (1993), Armstrong (1996), and Rochet and Choné (1999) for examples of monopoly firms who face consumers with multidimensional tastes. See Stole (2003) for a recent review of the competitive nonlinear pricing literature.

market for specialty coffee[3] near the University of Virginia (UVa), where coffee shops follow the common practice that larger drinks have lower per-ounce prices than smaller drinks. I specify a structural econometric utility model in which consumers vary in their vertical preferences for the shops' products and in their horizontal characteristics regarding their locations relative to the shops. The utility model is estimated using aggregate sales data and information on product and market characteristics. I use the utility estimates and cost data to compare consumers' benefits and the shops' costs for an additional ounce of a drink. This provides an estimate of the average distortion in product size for consumers who self-select into each purchasing option.

The estimated product allocation corresponds to some of the central predictions of second-degree price discrimination theory. Product sizes are close to efficient for the largest, most expensive espresso drinks; this mirrors the "no distortion at the top" result from theory. Distortions exist for the other specialty drinks – these products are generally too small – but inefficiencies decrease toward zero for products of increasing size and with the largest price-cost margins. Although the market studied here is of limited interest to policy makers, my empirical methods are applicable to a variety of markets. Further, the conditions that lead to distorted product designs are not restricted to traditional second-degree price discrimination settings. In many situations market power and asymmetric information matter for the design and pricing of consumer products or services.

Product design and efficiency have been central to the theoretical nonlinear pricing literature, but empirical research on second-degree price discrimination has primarily focused on other aspects of this pricing strategy. Many studies have asked whether observed price dispersion may be attributed to price discrimination.[4] Recently, other research has used structural econometric methods to analyze the welfare effects of observed and counterfactual market activity. Cohen (2005) and Verboven (2002) estimate markup patterns under equilibrium pricing assumptions and evaluate whether price discrimination occurs in paper towel and European auto markets, respectively. Miravete's (2002) research on telephone service pricing and Clerides' (2001) study of book publishing investigate the importance of unobserved consumer heterogeneity on product selection and welfare in monopoly markets. Ivaldi and Martimort (1994) answer similar questions with data from French farmers who purchased energy from competing firms. Leslie's (2004) study of

[3] "Specialty coffee" includes drip coffee made with high quality beans and espresso-based drinks such as lattes and cappuccinos. Specialty coffee shops (like Starbucks) generally post menus that include many drink types, with each type available in several sizes.

[4] See Shepard (1991), Borenstein and Rose (1994), and Busse and Rysman (2005).

monopoly pricing of theater tickets and Cohen (2005) compare the distribution of welfare under observed and counterfactual pricing schemes. Crawford and Shum's (forthcoming) analysis of monopoly cable TV firms is most similar to the present paper. They estimate the efficiency properties of cable TV packages under an assumption of optimal pricing and product design by cable firms.

This paper is organized as follows. In the next section I describe the UVa coffee market. I specify the empirical model of consumer utility in Section 3. Section 4 describes the data and how demand is identified. I also discuss what can be inferred about distortions from the data alone, without the utility model. The estimation strategy is described in the following section. In Section 6 I present the estimated utility parameters, and I report measures of product design distortions based on these estimates and the cost data. The final section concludes.

2 The Market for Specialty Coffee

In the late 1960s an interest in gourmet coffee began to grow in America, culminating in the recent specialty coffee boom.[5] Like many U.S. markets, the UVa's main campus ("central grounds") is served by several specialty coffee shops. During the sample period of February and March 2000, five firms operated nine coffee shops on campus or in the surrounding neighborhoods. The ARAMARK Corporation had four Greenberry's franchises; Espresso Corner, Starbucks, and Espresso Royale Caffe each operated one coffee shop on the UVa "Corner," a cluster of restaurants and shops adjacent to campus; and Higher Grounds had kiosks in the east and west cafeterias of the UVa Health Sciences complex.[6] The positions of the shops are illustrated in Figure 1.

Although nine shops served the market, it was never the case that all shops were open at once. The off-campus shops (Starbucks, Espresso Royale, and Espresso Corner) were open every day of the week from 7am or 8am until about midnight, but the other shops' schedules varied. Most variation in operating hours followed the opening and closing of the University facilities in which on-campus shops were located. This variation in schedules affected consumers' choice sets in a way that is useful for identifying the parameters of the empirical model. See Table 1 for information on the shops' operating hours.

[5] See Pendergrast (1999) for an overview of specialty coffee trends.

[6] The Greenberry's shops are the only franchises in the market. Espresso Corner and Higher Grounds are locally-based firms that have direct control over their coffee shops. Starbucks and Espresso Royale are national firms with local shop managers.

The coffee shops varied in the breadth of their menus, but all offered regular coffee.[7] Each shop employed nonlinear pricing. Firms with multiple coffee shops set one price schedule for all locations, and firms did not adjust their prices as competitors opened and closed. The sample period witnessed almost no price variation. Only Espresso Royale made any price changes, and this was just for one drink size in one product variety.[8] The pricing of two chains (Greenberry's and Higher Grounds) was regulated by the University because these firms operated their shops in UVa facilities. I was told by the owners of Higher Grounds that University oversight did not affect their pricing practices, but ARAMARK representatives reported that the University placed a binding ceiling on the prices charged at the Greenberry's shops. The prices at campus locations could be no higher than at an off-campus (unregulated) Greenberry's.

A distinguishing feature of the market is that most travel within the area is done on foot. This eases the study of the consumer population's location.[9] Parking in and near the market is largely limited to UVa employees and local residents, and the potential consumers' locations can be approximated by the distribution of the area's housing and the University's classroom and office space. I assume that the consumer population was limited to University employees and full-time students based in departments or facilities on campus.

3 The Empirical Model

3.1 Notation and simplifications

I assume that each consumer makes one purchasing decision during each time period, t. A period is defined as the portion of a day during which a constant number of coffee shops are open. This divides each day into three parts: "morning" (8am-2pm), "afternoon" (2pm-6pm), and "evening" (6pm-12am). The number of potential consumers during each period is N_t. Each person, i, buys

[7] Eight of the nine shops offered espresso-based drinks. Each shop sold food in addition to coffee, but I omit these products. I also ignore marketing practices that require dynamic decisions by consumers and firms. For example, a shop may sell a mug that allows consumers to return for discount refills. These practices are a small part of the market's activity.

[8] Espresso Royale offered a special on hot tea for the first six weeks of the sample period. Since I aggregate tea and drip coffee into one product line, the tea special appears in my data as a price reduction for 16 oz. servings of drip coffee.

[9] If a coffee shop is located in a strip mall, it is more difficult to characterize the roles of travel and distance in the market. See Thomadsen (2005) for a treatment of similar issues in a California fast food market.

one or zero cups of coffee during t. The decision not to buy specialty coffee is described as a choice of an outside good. This outside option includes actions like purchasing a soda, making coffee at home, or not consuming any beverage at all. All purchases of inside goods are limited to the coffee shops in the market. Purchasing options (the outside option and all specialty coffee products) are indexed by j, and those available during t are collected in the set J_t. For clarity, the outside option is assigned a separate abbreviation, o, in some equations below. The number of consumers who select option j during period t is $n_{jt} \geq 0$, and $N_t = \sum_{j \in J_t} n_{jt}$.

A wide variety of coffee and espresso drinks are available in the market, but I assume that all relevant beverages fall into one of three categories. The first possibility is "drip coffee," abbreviated as d in subscripts. This category includes all brewed coffee regardless of roasting darkness, country of origin, or caffeine content. Hot tea is also included in this drink class. The second category is "regular espresso," abbreviated r. Drinks in this category generally include shots of espresso and milk but no flavored syrups or whipped cream. The most popular regular espresso drinks are lattes and cappuccinos. The third category is espresso drinks containing syrups and whipped cream. Beverages with these ingredients are called "sweet espresso" and an s is used in subscripts. Each drink j is associated with exactly one of the product lines d, r, and s.

Across periods there may be differences in the market due to the identities of open shops, the location and number of consumers, or other market-wide common factors that shift the demand for coffee. I assume that these market-wide common factors are limited to changes in the value of the outside option across afternoon, evening, and weekend periods. Of the 21 periods in a week, 14 are identical in the consumers' geographic distribution, the choice set, and the value of the outside option. For convenience, I assign unique a "weekpart" index value to each of the 14 distinct combinations of market characteristics. Conditional on a set of prices, consumers make the same choices across periods (and weeks) in the same weekpart. The main advantage of imposing this structure is a reduction in the number of times predicted market shares must be calculated during estimation. For additional details on the rules that define weekparts, see Appendix A.

3.2 Consumers' decisions

Consumers differ in their taste for coffee and in their physical location; this heterogeneity has both vertical and horizontal components. Vertical preference heterogeneity enters the model through variation in the marginal utility from an additional ounce of a drink. Consumers who have a strong taste for one type of drink are not restricted to have similar tastes for drinks from other product lines. There are two sources of horizontal preference heterogeneity in the model. The first is variation in consumers' locations. The second source is unobserved idiosyncratic tastes for the offered products.

Utility maximization guides consumers' choices. Let U_{ijlt} be the utility from drink j to consumer i, who resides at location l during time t. Some attributes of drink j are at the shop level, but to minimize notational clutter I omit a shop index in the subscript. If consumer i does not purchase specialty coffee during t, he receives U_{iolt} in utility. Whenever i purchases the inside good j, the following inequalities must be satisfied:

$$\begin{aligned} U_{ijlt} &\geq U_{ij'lt} \;\; \forall\, j' \in J_t, \;\; \text{which includes} \\ U_{ijlt} &\geq U_{iolt}. \end{aligned} \tag{1}$$

The outside good is selected if U_{iolt} is larger than the utility from every inside good.

A consumer may purchase the specialty coffee j, which is of size q_j, has the price p_j, and is associated with product line $x \in \{d, r, s\}$. The consumer's utility from this action is affected by his personal taste (β_{xi}) for drinks in the same line as j; the distance (D_{jl}) from the consumer's location, l, to product j; and attributes (ξ_j) of j that are observed by agents in the market but are unobserved by the econometrician. These factors are combined to yield the conditional indirect utility function

$$U_{ijlt} = \beta_{xi} q_j^{\gamma_x} + \alpha p_j + \delta D_{jl} + \xi_j + \varepsilon_{ijlt}. \tag{2}$$

Consumer i's taste for product line x is specified as $\beta_{xi} = \beta_x \exp(\sigma_x v_{xi})$. The random components of i's taste for the three lines, v_{xi}, are drawn independently from the standard normal distribution. Thus, a consumer's taste for a drink from line x is the product of a population-wide median, β_x, and a unit-median deviation, $\exp(\sigma_x v_{xi})$. While a more general utility model would permit correlation across a consumer's values of v_{xi}, the data are insufficient to estimate these additional parameters and I assume they are equal to zero. The gross utility from q_j ounces of a drink from line x, $\beta_{xi} q_j^{\gamma_x}$, permits non-constant returns from quantity.[10] Additional random variation in U_{ijlt} enters through the error term ε_{ijlt}. This error captures idiosyncratic consumer heterogeneity in tastes, and it is uncorrelated with v_i. I assume that values of ε_{ijlt} are drawn independently from the type-I extreme

[10] A linear specification ($\gamma = 1$) would be inappropriate. If the marginal utility from an additional ounce is constant, price schedules are concave, and ε is omitted from the model, then all consumers with sufficiently high β_i to purchase a small coffee should prefer a large drink to the small one.

value distribution for each $ijlt$ combination. Additionally, the distribution of consumer distances to products, D_{jl}, is independent of v and ε.

Unobserved product characteristics are captured by the fixed effect ξ_j. I assume these effects are identical across all drinks within the same product line at a coffee shop.[11] A large value of ξ_j can explain why consumers would favor a drink with a relatively high price. A shop's values of ξ_j cover several product attributes. If a shop is popular because it has a comfortable seating area, this will be reflected in large values for each of the shop's ξ_js. If a shop has employees who are particularly good at making espresso drinks but tend to burn drip coffee, then this shop will have large values of ξ for both espresso lines and a smaller value for regular coffee. It is important to note that ξ and the gross utility from coffee quantity are additively separate.[12]

The utility to consumer i from not purchasing specialty coffee depends on time-varying market characteristics and the idiosyncratic taste shock ε_{iolt}. Let X_t be a vector of dummy variables indicating whether t is an afternoon, evening, and/or weekend period. Consumer i's utility from the outside good during t is

$$U_{iolt} = X_t\phi + \varepsilon_{iolt}, \tag{3}$$

where ϕ is a vector of parameters. Consumer i's unobserved taste for the outside good, ε_{iolt}, is drawn from the type-I extreme value distribution and is independent of all ε_{ijlt}.

Let θ represent a vector containing the parameters that determine the values of U_{ijlt} and U_{iolt}. The proportion of consumers who select a certain purchasing option depends on the observed attributes of all options in J_t, the parameter vector θ, and the distribution of observed and unobserved consumer characteristics. The distributions of D, v, and ε are denoted F_D, F_v, and F_ε, respectively. F_D depends on t because of changes to the distribution of consumers' locations. The proportion of the population that chooses option j during period t is

$$s_{jt}(\theta) = \int_{A_{jt}} dF_v(v)dF_\varepsilon(\varepsilon)dF_D(D;t), \tag{4}$$

where A_{jt} is the set of v, ε, and D values such that j is preferred to all other options in J_t. With N_t as the consumer population during t, the predicted market demand for j is $N_t s_{jt}$.

[11] For example, this product quality component is identical for 12 oz. and 16 oz. cups of drip coffee at Starbucks, but (possibly) different for 12 oz. lattes and drip coffees at the same store. The additional benefit of specifying a product-specific unobservable are likely to be minor. Bias in the estimation of α would arise only through variation in product quality within product lines.

[12] It may be desirable to allow β_i to vary across shops. However, this would either add more fixed effects or a non-additive random effect that is correlated with price.

Notice that the restrictions in (1) that determine (4) are essentially a set of incentive compatibility constraints and a participation constraint. These constraints play a central role in theoretical studies of price discrimination. However, unlike most theoretical studies, I cannot use (1) to formulate a set of relatively simple expressions that summarize the division of the market given in (4). There are two reasons for this. First, consumers are differentiated by a set of continuous variables but the set of products is discrete. This means that a consumer's purchasing decision cannot be characterized by a first-order condition that describes incentive compatibility. Second, the integral in (4) does not have a closed-form solution.

3.3 On the behavior of firms

Some, but not all, of the research on differentiated-products oligopoly markets employs equilibrium assumptions on firms' pricing behavior while estimating utility parameters.[13] The main advantages of these assumptions are: 1) improved precision in the demand estimates, and 2) the chance to impute marginal costs from first-order conditions on pricing despite a lack of cost data. I do not employ equilibrium pricing assumptions in my empirical model. I use the demand-side parameter estimates and production cost data to ask whether the product design predictions of second-degree price discrimination theory are observed in the market. Maintaining an assumption of equilibrium pricing may constrain the demand estimates to confirm the theory. In contrast, Crawford and Shum (forthcoming) assume optimal monopoly pricing and product design in order to estimate quality distortion magnitudes in cable TV services.

4 Data, Identification, and Preliminary Results

In this section I discuss the data and how they identify the empirical model. I describe in Appendix A how the data were collected and compiled. The section concludes with a discussion of the benefits and limits of investigating distortions directly in the price and cost data.

[13] Berry, Levinsohn and Pakes (1995), Cohen (2005), Petrin (2002), and Thomadsen (2005) use equilibrium assumptions during estimation. Leslie (2004) and Nevo (2000, 2001) estimate demand without these assumptions.

4.1 The data

4.1.1 Shop menus and sales data

I observe the UVa specialty coffee market for 51 days between February 1, 2000 and March 31, 2000 (all days excluding UVa's spring break). The total number of time periods, T, is 153. The sales data are aggregate figures for each purchasing option during each time period. For example, I observe that Espresso Corner sold n_{jt} units of its 16 oz. drip coffee during the morning of February 1^{st}. These sales data, together with the number of potential consumers in the market, are transformed into market shares. Table 2 contains summary information on market shares per period. On average, 3.27% of consumers purchased specialty coffee during each period, which means that there were about 800 specialty coffee sales. If no consumer purchased from a specialty coffee shop more than once per day, the percentage of consumers in the market who purchased coffee each day would be the sum of the morning, afternoon, and evening market shares (about 10%). Drip coffee products account for almost half of all sales, and regular espresso drinks were more popular than sweet espresso. A relatively large proportion of the consumer population purchased coffee in the morning, in part due to the length of the "morning" period (6 hours) relative to the "afternoon" (4 hours). A transaction is defined to be on the weekend if it occurs between Friday night and Sunday night, inclusive. All firms posted nonlinear prices for their products, and most of these prices exhibit quantity discounting.[14] Summary statistics on drink prices and prices per ounce are provided in Table 3.

4.1.2 Cost data

Using information on labor schedules, drink recipes, and materials expenses, I estimate the variable cost of serving each type of drink at each shop. There are two types of cost that might be called "marginal," so before proceeding I define the terms used to describe costs. The cost of one additional serving of drink j is called incremental cost, and the cost of increasing j's size by an ounce is called marginal cost. Marginal costs are the only cost information needed to evaluate whether a product is the efficient size, but incremental costs are necessary to describe profit margins.

I assume that product j's incremental cost is captured by the labor expense of preparing the drink plus the cost of the ingredients (coffee beans, milk, paper products, etc.). In this section I

[14] The only product line that featured (slightly) increasing prices per ounce is drip coffee at the Greenberry's shops. Not surprisingly, the largest drip coffee at Greenberry's was one of the least popular products in the market.

focus on the marginal cost estimates, since these costs are central to the welfare analysis below. I assume that a drink's labor cost is independent of product size but not product line. Details on labor costs data and analysis are in Appendix A. The main results are that labor's contribution to incremental cost is \$0.15 for drip coffee and \$0.38 for an espresso drink (regular or sweet).

Estimates of marginal costs come from data on ingredients expenses.[15] To keep my approach manageable, I select one popular drink from each of the Greenberry's drink lines and assume that all shops serve drinks that follow the Greenberry's recipes.[16] The drinks selected are: regular coffee with milk, a latte, and a "Milky Way" (essentially a latte with a several flavored syrups). The marginal cost for each type of drink is roughly constant. The mean cost is \$0.02 for an ounce of drip coffee, including accompanying condiments and paper products. The mean marginal costs for regular and sweet espresso drinks are \$0.04 and \$0.06, respectively. Additional information on the cost data is not reported because of confidentiality arrangements with the participating coffee shops. Summary statistics on incremental costs and markups are collected in Table 3. Average markups are larger than average incremental costs for all product lines, with price-cost margins generally largest for sweet espresso drinks.

4.1.3 Potential consumers and their locations

The number of consumers and their distribution throughout the market are important determinants of coffee shops' total sales. Using data from the Census and UVa's Registrar, Housing, and Facilities Management offices, I construct an empirical distribution of potential coffee shop customers for each weekpart. These distributions vary in the number of consumers in the market and their location. A simplifying assumption is to assign each consumer a "home" location for each period. I cannot account for the fact that a student could be in his dorm and several different academic buildings during a single period. The specification of home locations implies that a consumer who purchases specialty coffee makes a round trip between his home and the coffee shop.

The location data specify the probability that a randomly selected consumer has his home at one of L location points on or near campus, with $L = 239$. Off-campus locations are the centroids of 89 Census Bureau population blocks that surround campus. The remaining location points are

[15] All firms other than Espresso Royale provided cost data. I assign to Espresso Royale the market average of costs.

[16] With the help of a Greenberry's shop manager I obtained recipes that translate ingredients measurements into ounces of finished beverages.

centroids of University buildings. The shops on the UVa Corner are close to each other, so the average distance from a shop to its closest competitor is less than a fifth of a mile. The average distance between a shop and all other open shops is approximately half a mile. On average, a consumer has to travel 0.35 miles to reach the coffee shop closest to his home location.

4.2 Identification

While considering identification it is useful to keep in mind that coffee sales vary across shops within time periods as well as over time. The most important issue to address is identification of α, as there is minimal variation in the coffee shops' prices. Instead of "direct" price variation, I rely on changes in the price distribution as shops open and close. As the distribution of prices changes throughout the week, the share of consumers buying the inside goods changes as well. There is a correlation of -0.56 between the average price of products available during a weekpart and the percentage of consumers buying any specialty coffee drink. In Section 2 I argued that much of this variation in supply is exogenous, as it is due to the operating hours of the UVa facilities that house on-campus shops. In addition, changes in the distribution of consumer locations is analogous to the effect of opening and closing shops for a fixed consumer population.

The average popularity of each drink line identifies β_x, with $x \in \{d, r, s\}$. Departures from these averages for each shop allows estimation of the ξ_js, but the values of ξ_j within a product line are restricted to sum to zero across shops. Each product line's value of σ_x is identified by the substitution patterns across shops as the choice set changes. The parameters that determine the curvature of the utility function (the γ_xs) are identified by variation in the concavity of price schedules across shops. The disutility from travel (δ) is estimated using correlation between the distribution of consumer locations and the overall popularity of the inside goods. Changes in market-level sales across periods identifies the value of the outside option, ϕ.

4.3 Distortion evidence in the data

Before turning to the structural estimates, it is useful to note that the price and cost data provide some information on efficiency patterns in the market. In Table 4 I report statistics on the price differences (ΔP) between two adjacent (in size) drinks within a line and the cost difference (ΔC) for the same products. A consumer is unwilling to increase the size of his drink if the additional expense exceeds his value from the increase, but if the cost difference is relatively small it may be efficient for the consumer to buy a larger product. Thus, a large value of $\Delta P - \Delta C$ for a pair of products leads to inefficient allocations. Drip coffee and regular espresso have relatively large average values of $\Delta P - \Delta C$, at \$0.22 and \$0.23, respectively. The average value for sweet espresso is \$0.08. This implies that allocations are most efficient for sweet espresso, which I also find in the distortion estimates from the structural utility model. Further, the price and cost data suggest that the distortions for the largest espresso products are smaller than those for the smallest espresso drinks. This pattern is also supported in the distortion estimates below.

To appreciate the advantages of the structural estimates (and not stopping with Table 4), consider a pair of situations in which price and cost data alone could provide precise distortion measures. It would be ideal to observe firms that allow consumers to select any $q \in \mathbf{R}^+$, subject to the price function $p(q)$. With an observed marginal cost of c, a product menu's efficiency could be evaluated by comparing $p'(q)$ and c. Since each consumer would select q to equate marginal utility and $p'(q)$, unequal values of $p'(q)$ and c would reveal a distorted allocation immediately. If, instead, the firms offered a diverse but discrete menu with each product one ounce larger than the last, inference on distortions is more difficult but still feasible with price and cost data only. For a consumer who purchases exactly q ounces, the marginal price of the q^{th} ounce provides a lower bound on this increment's utility, and the marginal price of the $(q+1)^{st}$ ounce is an upper bound on utility for the next increment. Under a relatively weak assumption that marginal utility does not change too rapidly with q, the lower bound on utility from the q^{th} ounce can be compared to marginal cost to assess distortions. Also, the fine increments in product size suggest that consumers who select the same q are fairly homogenous in their utility from the next ounce. For the present paper, the structural model is necessary because the observed product menus are not so diverse. The statistics in Table 4 are informative about large changes in size only, and they do not account for differences among consumers who select the same drink. The parametric utility function in (2) provides a way to evaluate the marginal benefit from a small increase in product size. This is especially important for consumers who select the largest drinks, as I do not observe a marginal price they are unwilling to pay to receive more coffee. In addition, the structural utility model accommodates the taste heterogeneity of consumers who choose the same product in the UVa market.

5 Estimation

The utility parameters in (2) and (3) are estimated with maximum simulated likelihood (MSL).[17] Simulation is necessary because of the unobserved vertical preference heterogeneity in the consumer population, captured by the vector v. The log likelihood function is based on the multinomial distribution:

$$\mathcal{L}(\theta)=\sum_{t=1}^{T}\sum_{j\in J_t} n_{jt}\log[s_{jt}(\theta)]. \tag{5}$$

The main computational task in evaluating (5) is the calculation of the predicted market shares, s_{jt}. The parameterization of the utility function simplifies this process. Instead of calculating market shares for each product during each of the 153 time periods, shares are computed 28 times – the number of weekparts (14) × the number of price regimes (2).[18]

Further, consumer-to-shop distance information and the distribution of consumers within the market are available as data, so s_{jt} may be written as

$$s_{jt}(\theta) = w_{lt}\sum_{l=1}^{L} s_{jlt}(\theta).$$

s_{jlt} is the predicted share for option j within consumer location l, and w_{lt} is the proportion of consumers at l during t. The market share s_{jlt} requires integration over ε and v, but the former is possible analytically because of the functional form and independence assumptions on ε. Conditional on a consumer's location and value of v, the consumer selects product j with a multinomial logit (MNL) probability. The share s_{jlt} averages the MNL probabilities according to the distribution of v:

$$\begin{aligned} s_{jlt}(\theta) &= \int_{A_{jlt}} dF_v(v)dF_\varepsilon(\varepsilon) \\ &= \int_{\mathbf{R}^3}\left[\frac{\exp\{\overline{U}_{ijlt}(v)\}}{\sum_{j'}\exp\{\overline{U}_{ij'lt}(v)\}}\right]dF_v(v). \end{aligned} \tag{6}$$

In the first line of the expressions above, A_{jlt} is the set of (v,ε) values required by a consumer at location l to select option j. In the second line, integration is performed with respect to the variable v only, which has $\mathbf{R}^3$ as its support. The term $\overline{U}$ is a consumer's utility without the taste shock ε: $\overline{U} = U - \varepsilon$.

[17] I do not use GMM – the obvious alternative to MSL here – because my assumptions on the number and nature of the unobservables ξ allow direct estimation of these values as fixed effects.

[18] This means that period-specific factors (like weather) do not enter the choice problem, and, for example, each consumer finds the same solution to his decision problem on all Monday mornings during a given price regime.

I simulate s_{jlt} by drawing values of v, plugging these draws into $\overline{U}_{ijlt}$, and averaging over the resulting MNL probabilities. Because very few consumers actually purchase from a coffee shop, I employ importance sampling to over-sample from the right tail of v's support. This puts more "effort" into minimizing simulation error for the types of consumers who matter most for estimation. A detailed description of the likelihood function and related computational issues is in Appendix B.

The MSL estimate of the parameter vector, $\widehat{\theta}$, is a consistent estimator of θ only if both the numbers of simulations and observations go to infinity. However, there are two (related) reasons why this consistency result should not interfere with inference based on $\widehat{\theta}$. First, Börsch-Supan and Hajivassiliou (1993) show that MSL provides accurate parameter estimates in polychotomous choice problems even with a small number of simulations. Second, I use a number of simulations (1,000) that is not considered small for this type of problem. I construct the covariance matrix of $\widehat{\theta}$ with the robust "sandwich" formula (White, 1982) under the interpretation of (5) as a quasi-likelihood.[19]

6 Empirical Results

6.1 Parameter estimates and elasticities

Estimation results are reported in Table 5.[20] Standard errors are in parentheses; almost all of the parameter estimates are different from zero at high levels of significance. The estimates of price sensitivity (α) and the disutility from travel (δ) can be used to calculate the amount a person would pay ($\delta/\alpha = \$0.40$) to avoid adding 0.1 miles to each leg of a round trip between "home" and a coffee shop. If we assume that people typically walk at a pace of 20 minutes per mile, then these estimates imply that consumers would give up \$6 to avoid an hour of walking. This is close to the average hourly wage paid by the UVa coffee shops, \$6.70. The estimated values of β_x are difficult to interpret alone – a consumer's choices depend both on the population median β and his individual deviations from the median, which are influenced by σ. For example, the estimate of β_r (0.084) suggests that regular espresso drinks are unpopular with the median consumer. However, the relatively large value of σ_r (2.636) yields a sufficient number of consumers with high β_{ir} who are willing to buy a regular espresso product. The estimates of γ_x, which determine the curvature

[19] See McFadden and Train (2000) or Deb and Trivedi (2004) for similar examples of robust variance estimation within MSL.

[20] I estimate the model with an adjustment to the quantity variable. I divide the observed drink size by 12.

of utility, range from 0.334 to 0.415 and are all significantly different from one. This supports the argument in Section 3.2 against imposing linearity in the utility function.

Estimation restricts unobserved product line quality (ξ_j) to sum to zero for each line, so in maximizing (5) I omit the ξ_j values for Higher Grounds West. The fixed effects for this shop are determined by the estimates for the other shops, and the implied Higher Grounds West ξ_j values are given in italics. The fixed effects for the off-campus shops (Espresso Corner, Espresso Royale, and Starbucks) are large for all three product lines. Given the limitations of my data, it is not possible to determine whether this pattern is due to high drink quality or the design of the shops.[21] The stores on the UVa Corner control their decor, furniture, and entertainment (*e.g.*, live music). The other shops generally share space with facilities like cafeterias.

The outside option values during the afternoon and evening have the expected signs. Positive parameter values yield the relatively low market shares during these times of day (Table 2). A negative value of the outside option for weekends was not expected, since the market shares during these periods are fairly low. This parameter's value is likely due to the small number of shops that are open on weekends combined with a consumer population located farther away.[22]

Table 6 contains within-shop elasticity estimates, which are calculated for a 5% increase in price. An advantage of the random-coefficients utility specification is apparent. Cross-price elasticities within a product line are substantially larger than elasticities for drinks in different lines. The own-price elasticities may seem large in magnitude relative to those reported in other studies of multiproduct firms (*e.g.*, Nevo, 2000) given the size of the price-cost margins reported in Table 3. However, the low observed market shares imply that the outside option is popular, and consumers eagerly substitute toward it when a price increases. This interpretation in explored further in Section 6.3. Moreover, nonlinear pricing models with downward-binding incentive compatibility constraints have the property that a small price increase can lead to a large decrease in sales.[23] However, if prices increase for a pair of products adjacent in size, the same models predict a small

[21] It also may be due to my assumption that each consumer has a constant home. The shops on the UVa Corner are between a popular residential area and a group of academic buildings, and some of the shops' traffic is due to consumers walking between the two areas. I assume that customers make a less convenient round trip between the shops and a residential area or academic building.

[22] The unexpected result also might indicate a shortcoming of the consumer location data. I assume that consumers largely remain at home during the weekend, relatively far from the coffee shops. In fact, some consumers move through the market, and tourists may enter the market to visit campus.

[23] These issues are discussed in Rochet and Stole (2002) for a price-discriminating monopolist.

quantity decrease for the larger good. I therefore supplement the results in Table 6 by calculating the change in quantity when prices increase for all drinks in a firm's product line. A representative example from these calculations is the elasticity for a 12 oz. Higher Grounds coffee, which increases from -4.35 to -3.29 when the prices of adjacent products increase too.

6.2 Distortions in product design

I analyze the efficiency of product allocations by measuring the difference between the estimated marginal benefit and marginal cost from an additional ounce of a drink.[24] If this difference is positive for the consumers who purchase product j, then the consumers would purchase a larger product in a more efficient market. A monopolist facing consumers with a single dimension of private information will offer a product menu with positive differences between marginal benefit and cost for goods "low" within a product line and no distortion at the top of the line. The product characteristic distortions are tempered by competition. With limited competition, consumers self-select into purchases that entail reduced (but positive) inefficiencies. With sufficient competition consumers may obtain efficiently designed products.[25]

The distortion estimates are constructed as follows. Let $MB_j(\theta)$ be the average marginal benefit from an additional ounce of product j for consumers who choose this product. MB_j accounts for variation in the characteristics of consumers within and across time periods. I differentiate U_{ijlt} with respect to q_j and divide by $|\alpha|$ to obtain the marginal benefit (in dollars) of an increase in j's size for consumer i at location l:

$$MB_{ijlt}(\theta) = \frac{\gamma_x \beta_{xi} q_j^{(\gamma_x - 1)}}{|\alpha|}.$$

Product j is in drink line $x \in \{d, r, s\}$. Although no terms in MB_{ijlt} vary over time, several time-sensitive market characteristics (like consumer locations and the set of open shops) determine which consumers select option j during t. To obtain MB_j, I average MB_{ijlt} over consumer types (v), locations (l), and time periods within the sample (t). For products that are offered by the same firm at multiple coffee shops, averaging over shops is also necessary. The distortion $\Delta_j(\theta)$ is calculated as the difference between MB_j and the marginal cost of increasing j by an ounce.

[24] I do not consider distortions from the manipulation of product quality across product lines. Rochet and Choné (1995) and Armstrong (1996) suggest that one can observe distortions in one attribute (size) even if another attribute (quality) is unobserved by the econometrician.

[25] See Stole (1995), Rochet and Stole (2002), and Armstrong and Vickers (2001).

Details on the construction of MB_j and Δ_j are collected in Appendix B, including information on the distribution of Δ_j.

Estimated distortions are presented in Figure 2. The bar heights indicate the sizes of the distortions, and 95% confidence intervals are marked by vertical lines. The distortion values are given in pennies. The estimated distortions for most espresso products are positive, and they conform to results from price discrimination theory. The firms' sweet espresso lines, the top products in price and usually in markup, feature distortions that decrease with cup size. The mean distortion (across firms) is 3.66 cents for the smallest sweet espresso drinks, while the mean distortion is 0.16 cents for the largest drinks in this line. This mirrors the "no distortion at the top" result from theory. Two firms have negative Δ_j values for their largest sweet espresso drinks, but both negative values are small in magnitude and closer to zero than the distortions for the next-smallest drinks.

Each shop's distortions of regular espresso are larger than the corresponding distortions for sweet espresso. In interpreting Figure 2, it is important to remember that there is not a clear ranking of all products, as there might be if the shops sold many sizes in a single product line. The pattern of distortions is consistent with nonlinear pricing theory, which predicts that lower-profit products (regular espresso) are degraded to become poorer substitutes for higher-profit items (sweet espresso). However, the small cross-price elasticities for drinks in the two espresso lines present a puzzle as to why a shop might distort the size of the largest regular espresso drink. The likely answer is in the specification of the utility model. I constrain tastes across product lines to be uncorrelated. The estimated distortion pattern would be reasonable if consumers' tastes for regular and sweet espresso are positively correlated. This idea is explored further in the next subsection.[26]

Finally, the drip coffee lines have positive distortions that generally increase gradually with cup size. The mean benefit-cost difference is 2.3 cents for the smallest drip coffees, and the mean

[26] An alternative view is that the distortions may be in place because of the congestion externality that espresso customers impose on each other due to the lengthy drink preparation. As a customer waits in line, this person's surplus is reduced without being captured by the shop. The firms may use their menus to insure that the highly valued sweet espresso customers are ordering their drinks with maximal surplus left to extract. Relatively high prices and distorted characteristics may reduce demand by potential regular espresso customers while indirectly increasing the chance that a sweet espresso customer will purchase from a shop. It is not clear which circumstances (if any) support this menu design strategy as optimal. Joint consideration of negative externalities and nonlinear pricing is not a part of the existing literature, and an innovation on this front is beyond this paper's scope.

difference is 2.6 cents for the largest drinks. These relatively small distortions are similar to the predictions of Rochet and Stole (2002), who find efficiency at the top *and* bottom of a unidimensional product menu in monopoly and duopoly markets for consumers with private vertical and horizontal tastes. However, it is important to note that Rochet and Stole consider a model in which vertical and horizontal taste heterogeneity are each unidimensional. Their qualitative results may change if the model included the menu and consumer characteristics described in Sections 2 and 3. Alternatively, drip coffee is in closer competition with the outside good than are the espresso drinks. This may induce the specialty coffee shops to offer drip products close to the efficient size.

6.3 Robustness tests

I verify the robustness of my results by re-estimating the model under different functional form assumptions for utility and with alternative data. These results are reported in Table 7. The first column ("Original Model" – Specification A) reproduces estimates from Tables 5 and 6. Specifications B and C employ alternative assumptions about utility. In B, I impose a correlation of 0.5 between consumers' tastes for regular and sweet espresso, with all other aspects of the utility function unchanged from A. There is a substantial increase in the estimated cross-price elasticities between the product lines, but no appreciable change in the estimated distortions. As noted above, this correlation might explain the positive distortions at the top of the regular espresso lines. In Specification C, I assume that vertical tastes are $\beta_{ix} = \beta_x v_i$, where v_i is distributed $\chi^2_{(2)}$. The distortion estimates in A and C are similar. In Specification D, I estimate the same model as in A, but with half the consumer population. This assumption acknowledges that many residents may be completely uninterested in purchasing coffee. The distortion estimates in D are again similar to those in A, but the reported demand elasticity increases to -3.90. A smaller total population and unchanged aggregate sales reduce the popularity of the outside option, and there is less substitution toward it. Finally, in Specification E I exclude data from Starbucks and Espresso Corner since I relied on surveys to construct some data for these shops.[27] The important features of the distortion estimates are unchanged in E, with minor differences in the slopes of the drip coffee and sweet espresso estimates. Demand is more elastic in E due to the (intentional) misspecification.

[27] To estimate Specification E, I include a catch-all fixed effect for consumers who travel to a location between Starbucks and Espresso Royale instead of purchasing from one of the remaining shops. This accounts for the differential impact of excluding these shops on consumers in different parts of the market. Additionally, I fix the values of γ at those in Specification A to simplify estimation.

In omitting sales at two popular shops, the outside option's relative popularity increases.

7 Conclusions

In this paper I conduct an empirical examination of nonlinear pricing in an oligopoly market. My primary focus is on distortions in product allocations, which are central to the theoretical literature but have received little attention in empirical work. I use data on sales, product attributes, and consumer locations to estimate a structural utility function. My model includes separate vertical taste heterogeneity for three product types (drip coffee, regular espresso, and sweet espresso), and horizontal heterogeneity through differences in consumer locations. The estimated utility model, combined with the cost data, is used to calculate the average allocative inefficiency for each product. The estimated distortion patterns match some of the predictions from the theoretical literature. As in Rochet and Stole (2002), distortions are close to zero at the top of the product menu (largest sweet espresso) and bottom (smallest drip coffee), with larger positive benefit-cost margins in between. The finding that distortions decrease toward zero at the top is the first empirical support for a pattern predicted by Mussa and Rosen (1977) and Maskin and Riley (1984). Since the utility model and distortion patterns are estimated without assumptions on firm behavior, the estimated distortions are not restricted to correspond to predictions from theory.

Many markets may be examined using similar methods. Most items in supermarkets are available in multiple package sizes with nonlinear pricing. Banking services like loans and checking accounts are designed with a several different "prices" (interest rates and other requirements). Consumer electronics items like computers and televisions are sold with different size or quality attributes. In these markets there is a socially efficient set of product characteristics for each consumer, but market power and information asymmetries may lead to distortions. My methods can be employed to judge whether allocative inefficiencies are large. It is not necessary to obtain marginal cost data to perform this analysis. Under assumptions regarding equilibrium conduct by firms, marginal cost information can be imputed within an estimation procedure.

My paper's empirical methods may lead to different policy conclusions than alternative analyses. Consider the sales of the smallest cups of drip coffee. The raw data indicate that there are large markups for these products. However, the estimated distortions show that these drinks are close to the efficient size for the consumers who purchase them. Intervention in the market will not increase substantially the total surplus from serving these consumers, although it may affect the distribution of surplus. Welfare losses may be limited to consumers excluded from the market at the observed prices.

A Appendix: The Data

A.1 Sales data

Methods for collecting sales data varied across the firms. This variation was largely due to differences in the firms' preferences and costs for revealing their sales data to me. The ARAMARK Corporation, which operated the four Greenberry's shops in the market, provided cash register tapes for each shop during the full sample period. These tapes include information (date, time, and drink) about each transaction made.

Espresso Corner provided daily cash register tapes that include the number of sales in each drink category. These tapes were run at approximately 1pm, 6pm, and midnight on most days. The end-of-day tapes are available for each day in the sample period, but some mid-day tapes are missing or were run at odd times. To predict sales during the each part of a day, I run a regression of mid-day sales on a constant and the deviation (in minutes) of the tape time from the nearest period cut-off time defined in Section 4.

The owners of Higher Grounds provided their daily revenue figures for each of two shops in the market, plus a number of cash register tapes that track revenue during the day. I use these tapes to construct data on the sales distribution throughout the day. To construct data on menu choices, I observed 381 transactions at the east cafeteria location and 198 transactions at the west cafeteria shop. Because of data limitations I assume that the observed distribution of menu selections is constant throughout the sample period.

Neither Starbucks nor Espresso Royale Caffe (ERC) were willing to share data on total sales, so I employed a team of volunteers to count the number of people entering each shop during various parts of the day and days of the week. These volunteers counted Starbucks patrons for 86 hours and ERC patrons for 87 hours. Approximately 3,200 customers were counted at each shop. Additionally, 363 Starbucks customers and 392 ERC customers were polled about their purchases. The traffic counts and poll results are used to construct sales data for Starbucks and ERC. Starbucks provided data on the share of store sales in each drink category, so the Starbucks poll results are needed only to estimate the number of counted patrons who did not buy a drink. The sales data for ERC are constructed similarly but rely on polling results rather than firm-provided figures for menu selection data. I assume that Starbucks' distribution of menu selections is stable within the entire sample period; for ERC I assume a stable distribution within each price regime.

A.2 Consumer location data

The data on consumer locations are points in space to which I assign a use or uses (*e.g.*, classroom or classroom and office space) and probability weights. Each location point can be classified as a place to work or sleep. I assume that the market is flat and Euclidean distance is used to measure travel between locations. The market's geographic boundaries are major roads or areas without much housing. It would be inconvenient to travel by foot into the market from locations beyond these boundaries.

Data on places to sleep ("beds") are constructed using the 1990 Census,[28] projections of population growth from the Census Bureau, and information on dormitory space from UVa's Housing Services offices. There are 57 points in the market that are associated with dormitory buildings. Each dormitory location has between 20 and 800 beds. The number of off-campus beds in surrounding neighborhoods is compiled using population data at the census block level. I assume that off-campus beds are occupied by UVa students only. The number of University students is large enough that all of the available dorm rooms and all of the housing in surrounding neighborhoods are necessary to provide beds for the student population.

Data on working (non-bed) locations come from the UVa Facilities Management Department. For each on-campus building I observe the number of designated work stations and the typical use for each station. I treat classroom and non-classroom stations separately, and I use the building's number of stations of each type to determine the probability that a student is in class there or an employee works in the building.

The student population is the group of 16,491 people based in departments or schools that are on UVa's central campus. All students are assumed to be in class, in a library, or in bed. The numbers of students in class or in a library are constructed with data from the UVa Registrar and library traffic reports. The number of students in bed during any period is the residual. I assume that the number of students in class is equal to the cumulative enrollment during a weekpart.[29] Table A1 displays the cumulative enrollment for each weekpart. If the number of enrollments is larger than the total student population, I assume that the entire student body is in class.

The University employee population comprises all full- and part-time workers who are based

[28] The block-level results of the 2000 decennial census were not available when I compiled the data for this research.

[29] This will over-count consumers but it seems more appropriate than using the number of enrolled students during a particular hour. For example, the total number of enrollments on Monday mornings is almost 19,000, but the maximum number of students in class during any hour is 4,000.

on the main campus. Employees are classified as salaried and non-salaried, and hospital and non-hospital. I assume that the 9,463 non-hospital salaried workers are in the market during weekday mornings and afternoons, as are three-quarters of non-hospital non-salaried workers. The remainder of non-hospital employees work during weekday evenings and on weekends. I make a similar assumption about the schedules of the 1,546 hospital employees. Two-thirds of these workers are in the market during weekdays, and the others work at night and on the weekend. The numbers of non-classroom work stations in UVa buildings are used to construct the distribution of UVa employees in the market. Since I assume that the housing surrounding campus serves students only, University employees exit the market when they are not at work.

A.3 Cost data

A.3.1 Ingredients costs

I have precise data on drink ingredient costs for Greenberry's, Higher Grounds, and Espresso Corner. Starbucks provided some data on ingredients costs but at a lower level of detail. Espresso Royale declined to provide any cost data. My approach to marginal cost requires a measure of a one ounce increase in drink size. There are three important components of this cost: the drink ingredients themselves, paper products, and condiments. I review the specific assumptions about the components here.

The role of drink ingredients is simplest in the case of drip coffee – an additional ounce of a drink means that more coffee is poured. Espresso products are more difficult because of the discreteness in espresso shots, plus the proportion of espresso in a drink can vary within a product line. To deal with these issues, I construct the marginal cost of an espresso drink under the assumption that all ingredients remain in the same proportions as drink size increases. Because espresso drinks may have different ratios of espresso to steamed milk, the per-ounce costs may differ across drinks in a shop's product line. The differences within a shop's espresso lines are typically less than a penny

Next, consider the role of paper products. The only non-negligible paper expenses are for drink cups and lids. The main hurdle is that cup and lid costs are observed for discrete sizes only. When a coffee shop serves 12, 16, and 20 oz. drinks, it is difficult to calculate the extra expense of changing the 12 oz. drink to 13 oz. Fortunately, cup and lid expenses increase gradually through drink sizes, so the impact of paper products on marginal cost is small. I calculate the average expense per shop for moving between observed drink sizes (*e.g.*, from 12 oz. to 16 oz.), and then divide by the size of the observed difference. Computed this way, paper's contribution to marginal costs is about half a penny.

Finally, condiments such as milk, sugar, and artificial sweeteners are the smallest part of marginal costs. All condiments other than milk are of negligible expense, so I ignore them while calculating costs. I assume that milk is added to drip coffee only, at a rate of 1 oz. of milk per 8 oz. of coffee. Milk is then a minor part of one additional ounce of a drink, and contributes about a fifth of a cent to marginal cost.

A.3.2 Labor costs

Labor schedule and wage data are available for the two most popular Greenberry's shops, both Higher Grounds locations, and Espresso Corner. I use this information along with the sales data to construct estimates of the labor cost of serving drip coffee and espresso drinks. Each shop's labor schedules are constant for each weekpart within the sample period. I interpret this regularity to mean that the shops' managers set their labor schedules based on expected sales during the different parts of the week. Ideally, I would like to regress labor expenditures on expected sales, but the expected values are not available as data. Average realized sales replace expected sales in the analysis. I consider sales data on the average numbers of drip coffee and espresso transactions ($trans_{dt}$ and $trans_{et}$) and the average numbers of ounces served of drip coffee and espresso ($ounces_{dt}$ and $ounces_{et}$). I aggregate sales of regular and sweet espresso drinks for my analysis of labor costs.

I run several regressions of labor costs on sales and a set of shop-level dummy variables.[30] Results are collected in Table A2. The regressions yield an average value of $17 for shop dummy coefficients. This should be interpreted as the average expense to a coffee shop from having employees open or close registers for the day, mop the floors, and perform other tasks not immediately related to selling an additional cup of coffee. The sales coefficient estimates are statistically significant only when either the *trans* or *ounces* variables are included (not both). This is almost certainly due to multicollinearity among the variables.

Among the results reported on Table A2, I use those of Specification A as my estimates of labor costs. These estimates indicate that the per-drink labor costs are $0.15 for drip coffee and $0.38 for espresso products. Specification A is reasonable if the most important labor requirements are independent of serving size (*e.g.*, taking the order, ringing up the sale, and preparing the basic

[30] The main appeal in using labor costs instead of hours is that the former accounts for differences in worker productivity and the corresponding variation in wages.

materials and ingredients). Using only the *trans* variables means that the labor expense of serving an 8 oz. drink is the same as serving a 20 oz. drink. This seems preferable to the assumption inherent in the models with the *ounces* variables, that the labor expense of a 20 oz. drink is more than twice that of an 8 oz. drink.

B Appendix: Estimation Details

B.1 Utility parameter estimates

The likelihood function is based on the multinomial distribution. Consider an experiment that is repeated independently for T time periods. During each period, t, N_t independent draws are made from a collection of $|J_t|$ cells. The probability that a draw comes from the j^{th} cell is s_{jt}, and s_t is the vector of $|J_t|$ probabilities for t. The outcome is a vector $n_t = (n_{1t}, .., n_{jt}, ...n_{|J_t|t})$, where n_{jt} is the number of times option j is selected. The joint probability of observing the sequence of outcomes $\{n_t\}_{t=1}^{T}$ given the vectors of probabilities $\{s_t\}_{t=1}^{T}$ is the product of T mass functions of the multinomial distribution:

$$L(\theta) = \prod_{t=1}^{T} \left(N_t! \prod_{j \in J_t} \frac{s_{jt}(\theta)^{n_{jt}}}{n_{jt}!} \right). \tag{7}$$

The outcomes (n_t) and cell probabilities (s_t) correspond to the observed purchasing decisions and predicted market shares of the empirical model. The predicted market shares are the only parts of (7) that depend on the parameter vector θ. After taking the natural logarithm of $L(\theta)$ and eliminating terms unaffected by θ, I obtain the log-likelihood function (5). Let $\mathcal{L}_{jt} = n_{jt} \log[s_{jt}(\theta)]$ be the likelihood contribution of the j^{th} purchasing option during the t^{th} period.

The main computational task in evaluating (5) is the calculation of predicted market shares for each product during each period. In Section 5 I show how the distributions of D and ε may be handled during estimation, but v requires simulation. The simulation problem reduces to calculating MNL probabilities for the location-specific market shares in (6). Simulation of v involves taking three independent draws from $U[0, 1]$ to obtain the vector $z = (z_1, z_2, z_3)$, and then evaluating $\Phi^{-1}(z_i)$, where Φ is the univariate standard normal distribution function. The one-to-one relationship between the distributions of v and z implies that (6) can be written in the following simplified form:

$$s_{jlt}(\theta) = \int_{\mathbf{U}^3} g(z; \theta) d\Upsilon(z). \tag{8}$$

Υ is the distribution function of z and $\mathbf{U}^3 = [0,1] \times [0,1] \times [0,1]$ is z's support. The function g replaces the MNL probability in (6) and incorporates the mapping of z to v. A straightforward way to simulate s_{jlt} is to draw R values of z from Υ and take the average of a sequence of $g(z^r;\theta)$ values:

$$\widetilde{s}_{jlt}(\theta) = \frac{1}{R}\sum_{r=1}^{R} g(z^r;\theta).$$

While $\widetilde{s}_{jlt}$ is an unbiased simulator of s_{jlt}, this approach is not optimal.

I construct an alternative simulator, $\widehat{s}_{jlt}$, that uses a larger percentage of draws from sections of $\mathbf{U}^3$ where at least one value of z_i is relatively large. This is a form of importance sampling.[31] The support of each entry in z is split into two regions – a low section and a high section. Let $\widehat{z}_i \in (0,1)$ be the value that divides the support of z_i, so that all values of z_i are drawn from either $[0,\widehat{z}_i]$ or $[\widehat{z}_i,1]$. The partition of $\mathbf{U}^3$ includes all combinations of these low and high regions for each entry in z; there are 8 such combinations. These regions are denoted $\mathbf{U}^3_k$.

This partition of $\mathbf{U}^3$ can be used in rewriting (8) as a sum of conditional probabilities:

$$s_{jlt}(\theta) = \sum_{k=1}^{8} \Pr(z \in \mathbf{U}^3_k) \int_{\mathbf{U}^3_k} \frac{g(z;\theta)d\Upsilon(z)}{\Pr(z \in \mathbf{U}^3_k)}.$$

To construct the simulator $\widehat{s}_{jlt}$, I draw R_k times from each region $\mathbf{U}^3_k$, with $\sum_k R_k = R$. These draws are used to generate (separate) simulators of each integral in the sum above. To simplify notation, replace $\Pr(z \in \mathbf{U}^3_k)$ with ρ_k and write:

$$\widehat{s}_{jlt}(\theta) = \sum_{k=1}^{8} \frac{\rho_k}{R_k}\left(\sum_{r=1}^{R_k} g(z^r;\theta)\right).$$

To implement this simulator, I chose $\widehat{z}_i = 0.7$ for all i and $R_k = 125$ for all k. The probabilities ρ_k follow immediately from the vector of thresholds, $\widehat{z}$. For example, a vector z drawn from Υ has all entries below 0.7 with probability 0.343, and there is a probability of 0.063 that the first two entries are above 0.7 and the third is below 0.7. Whereas the simple simulator $\widetilde{s}_{jlt}$ has approximately 0.343 and 0.063 of its draws in these regions, respectively, the simulator $\widehat{s}_{jlt}$ uses the same number of draws in each region.

To construct the covariance matrix of $\widehat{\theta}$, I follow the robust "sandwich" procedure in White (1982). Let M represent the total number of observations, summed over time periods and purchasing options. I construct the matrix $\widehat{\Omega}$ as the outer product of $\mathcal{L}$'s gradient:

$$\widehat{\Omega} = \frac{1}{M}\sum_{t=1}^{T}\sum_{j\in J_t} \frac{\partial \mathcal{L}_{jt}(\widehat{\theta})}{\partial\theta}\frac{\partial \mathcal{L}_{jt}(\widehat{\theta})}{\partial\theta'}.$$

[31] Berry, Levinsohn, and Pakes (1995) use a similar procedure. See Stern (1997) for a review of simulation methods.

$\widehat{H}$ is an estimate of $\mathcal{L}$'s Hessian matrix:

$$\widehat{H} = -\frac{1}{M}\sum_{t=1}^{T}\sum_{j\in J_t}\frac{\partial^2\mathcal{L}_{jt}(\widehat{\theta})}{\partial\theta\partial\theta'}.$$

The sandwich estimator of $\widehat{\theta}$'s covariance matrix is $V(\widehat{\theta}) = \widehat{H}^{-1}\widehat{\Omega}\widehat{H}^{-1}$. I use $V(\widehat{\theta})$ to compute the standard errors in Table 5, and $V(\widehat{\theta})$ is an input for calculating the confidence intervals in Figure 2.

B.2 Distortion estimates

In this section I provide details on how to construct $MB_j(\theta)$ from the term $MB_{ijlt}(\theta)$ defined in Section 6.2. I first consider the average marginal benefit to all consumers at l who purchase j during period t. The average marginal benefit to these consumers, conditional on purchasing j, is

$$MB_{jlt}(\theta) = \frac{\int_{A_{jlt}} MB_{ijlt}(\theta)dF_v(v)dF_\varepsilon(\varepsilon)}{\int_{A_{jlt}} dF_v(v)dF_\varepsilon(\varepsilon)}. \tag{9}$$

In evaluating the integrals in (9) I employ the same simplifications and numerical techniques as in the estimation routine. MB_{jt} is calculated as a weighted average of MB_{jlt} terms over location points, just as in the case of market shares. The final expression for MB_j is an expected sales-weighted average over values of MB_{jt}:

$$MB_j(\theta) = \frac{1}{\sum_t N_t s_{jt}}\left(\sum_{t=1}^{T} N_t s_{jt} MB_{jt}(\theta)\right).$$

This expression for purchasing option j is the appropriate measure of marginal benefit if the firm offering j operates only one shop. For firms with multiple shops, I calculate the average benefit (across shops and time periods) from adjusting a drink's size at all shops. This corresponds to the practice of multi-shop firms of offering the same menu at each shop.

The distortion for product j, $\Delta_j(\theta)$, is the difference between MB_j and the marginal cost of increasing j's size by an ounce. Since I have information on marginal costs, MB_j is the only component of Δ_j that depends on the estimated parameter vector. Δ_j is a statistic with a variance that is determined entirely by the distribution of MB_j. I compute the variances of the Δ_js with a bootstrapping procedure. I draw B times from the sampling distribution of $\widehat{\theta}$ to generate the sequence $\{\theta^b\}_{b=1}^B$. The variance of Δ_j is calculated as the average of squared differences between $\Delta_j(\theta^b)$ and $\Delta_j(\widehat{\theta})$:

$$\text{Var}[\Delta_j(\theta)] = \frac{1}{B}\sum_{b=1}^{B}[\Delta_j(\theta^b) - \Delta_j(\widehat{\theta})]^2.$$

$B = 1{,}000$ is used for the reported results.

References

[1] Armstrong, M. "Multiproduct Nonlinear Pricing." *Econometrica*, Vol. 64 (1996), pp. 51-74.

[2] Armstrong, M. and Vickers, J. "Competitive Price Discrimination." *RAND Journal of Economics*, Vol. 32 (2002), pp. 579-605.

[3] Berry, S., Levinsohn, J., and Pakes, A. "Automobile Prices in Market Equilibrium." *Econometrica*, Vol. 63 (1995), pp. 841-890.

[4] Borenstein, S. and Rose, N. "Competition and Price Dispersion in the U.S. Airline Industry." *Journal of Political Economy*, Vol. 102 (1994), pp. 653-683.

[5] Börsch-Supan, A. and Hajivassiliou, V. "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models." *Journal of Econometrics*, Vol. 58 (1993), pp. 347-368.

[6] Busse, M. and Rysman, M. "Competition and Price Discrimination in Yellow Pages Advertising." *RAND Journal of Economics*, Vol. 36 (2005), pp. 378-390.

[7] Clerides, S. "Product Selection with Multi-Peaked Preferences in Book Publishing." University of Cyprus working paper, 2001.

[8] Cohen, A. "Package Size and Price Discrimination in the Paper Towel Market." Federal Reserve Board of Governors working paper, 2005.

[9] Deb, P. and Trivedi, P.K. "Specification and Simulated Likelihood Estimation of a Non-normal Treatment-Outcome Model with Selection: Application to Health Care Utilization." Indiana University working paper, 2004.

[10] Ivadi, M. and Martimort, D. "Competition under Nonlinear Pricing." *Annales d'Economie et de Statistique*, Vol. 34 (1994), pp. 71-114.

[11] Leslie, P. "Price Discrimination in Broadway Theatre." *RAND Journal of Economics*, Vol. 35 (2004), pp. 520-541.

[12] Maskin, E. and Riley, J. "Monopoly with Incomplete Information." *RAND Journal of Economics*, Vol. 15 (1984), pp. 171-196.

[13] McFadden, D. and Train, K. "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics*, Vol. 15 (2000), pp. 447-470.

[14] Miravete, E. "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans." *Review of Economic Studies*, Vol. 69 (2002), pp. 943-971.

[15] Mussa, M. and Rosen, S. "Monopoly and Product Quality." *Journal of Economic Theory*, Vol. 18, pp. 301-317.

[16] Nevo, A. "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." *RAND Journal of Economics*, Vol. 31 (2000), pp. 395-421.

[17] Nevo, A. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, Vol. 69 (2001), pp. 307-342.

[18] Pendergrast, M. *Uncommon Grounds*. New York: Basic Books, 1999.

[19] Petrin, A. "Quantifying the Benefits of New Products: the Case of the Minivan." *Journal of Political Economy*, Vol. 110 (2002), pp. 705-729.

[20] Rochet, J.-C., and Choné, P. "Ironing, Sweeping, and Multidimensional Screening." *Econometrica*, Vol. 66 (1998), pp. 783-826.

[21] Rochet, J.-C., and Stole, L. "Nonlinear Pricing with Random Participation Constraints." *Review of Economic Studies*, Vol. 69 (2002), pp. 277-311.

[22] Shepard, A. "Price Discrimination and Retail Configuration." *Journal of Political Economy*, Vol. 99 (1991), pp. 30-53.

[23] Shum, M. and Crawford, G. "Monopoly Product Degradation in Cable Television." *Journal of Law and Economics*, forthcoming.

[24] Stern, S. "Simulation-Based Estimation." *Journal of Economic Literature*, Vol. 35 (1997), pp. 2006-2039.

[25] Stole, L. "Nonlinear Prices and Oligopoly." *Journal of Economics and Management Strategy*, Vol. 4 (1995), pp. 529-562.

[26] Stole, L. "Price Discrimination in Competitive Environments." University of Chicago GSB working paper, 2003.

[27] Thomadsen, R. "The Effect of Ownership Structure on Prices in Geographically Differentiated Industries." *RAND Journal of Economics*, Vol. 36 (2005), pp. 908-929.

[28] Verboven, F. "Quality-Based Price Discrimination and Tax Incidence: Evidence from Gasoline and Diesel Cars." *RAND Journal of Economics*, Vol. 33 (2002), pp. 275-297.

[29] Wilson, R. *Nonlinear Pricing.* New York: Oxford University Press, 1993.

[30] White, H. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica*, Vol. 50 (1982), pp. 1-25.

TABLE 1

Coffee Shop Schedules

	Mon – Wed	Thurs	Fri	Sat	Sun
Morning (8am – 2pm)	GA, GB, GP, HGE, HGW	GA, GB, GP, HGE, HGW	GA, GB, GP, HGE, HGW	GB, GP, HGE	GB, GP, HGE
Afternoon (2pm – 6pm)	GA, GB, GP, HGE	GA, GB, GP, HGE	GA, GB, GP, HGE	GB, GP	GA, GB, GP
Evening (6pm – 12am)	GA, HGE	GA, GT, HGE	GT, HGE	GT	GA

The shop abbreviations: Greenberry's Alderman (GA), Greenberry's Bookstore (GB), Greenberry's Poolside (GP), Greenberry's Tuttle (GT), Higher Grounds East (HGE), and Higher Grounds West (HGW). Starbucks, Espresso Corner, and Espresso Royale were open during all time periods, so these shops are excluded from the table.

TABLE 2

Mean Market Shares

All specialty coffee	3.27 (1.45)	Morning (8am - 2pm)	5.10 (0.98)
Drip coffee	1.57 (0.90)	Afternoon (2pm - 6pm)	2.26 (0.30)
Regular espresso	0.95 (0.27)	Evening (6pm - 12am)	2.43 (0.32)
Sweet espresso	0.74 (0.30)	Weekend	2.68 (0.64)

The means are calculated by summing sales across all shops during a period, and then dividing by the market population. Standard deviations are in parentheses.

TABLE 3

Drink Prices and Costs (Dollars)

Product Line	Price	Price/Ounce Smallest	Price/Ounce Largest	Incremental Cost	Price-Cost Margin
Drip Coffee	1.41 (0.31)	0.120 (0.03)	0.094 (0.01)	0.41 (0.08)	1.00 (0.29)
Regular Espresso	2.49 (0.53)	0.200 (0.01)	0.155 (0.01)	0.96 (0.20)	1.53 (0.43)
Sweet Espresso	3.04 (0.51)	0.260 (0.03)	0.185 (0.02)	1.34 (0.32)	1.70 (0.37)

Standard deviations are in parentheses.

TABLE 4

Investigating Distortions Directly from Data (Dollars)

	Mean $\Delta P - \Delta C$		
Product Line	All Drinks	Smallest	Largest
Drip Coffee	0.22	0.16	0.29
Regular Espresso	0.23	0.29	0.21
Sweet Espresso	0.08	0.13	0.06

TABLE 5

Utility Parameter Estimates

Price (α) :	-4.509 (0.064)	Distance (δ) :	-1.802 (0.081)

	Drip Coffee	Regular Espresso	Sweet Espresso
Median utility (β)	0.388 (0.050)	0.084 (0.009)	1.213 (0.107)
Preference variation (σ)	1.643 (0.072)	2.636 (0.056)	1.260 (0.046)
Utility curvature (γ)	0.415 (0.071)	0.360 (0.014)	0.334 (0.008)
Fixed Effects (ξ)			
GB – Alderman	-1.049 (0.142)	-0.915 (0.175)	-1.108 (0.210)
GB – Bookstore	-1.573 (0.088)	-2.021 (0.087)	-1.494 (0.200)
GB – Poolside	-3.260 (0.058)	N/A[1]	N/A[1]
GB – Tuttle	-2.416 (0.032)	-0.720 (0.303)	-3.277 (0.243)
Espresso Corner	2.233 (0.154)	1.547 (0.107)	2.718 (0.120)
Starbucks	1.963 (0.095)	3.281 (0.161)	3.018 (0.138)
Espresso Royale	1.686 (0.043)	2.011 (0.097)	2.059 (0.309)
HG – East Cafeteria	1.813 (0.115)	-0.796 (0.292)	0.299 (0.146)
HG – West Cafeteria	*0.603* (*0.137*)	*-2.387* (*0.121*)	*-2.215* (*0.417*)

	Afternoon	Night	Weekend
Outside Options (ϕ)	1.650 (0.177)	0.651 (0.137)	-0.926 (0.252)

Standard Errors are in parentheses. Abbreviations: GB = Greenberry's, HG = Higher Grounds
1: Greenberry's Poolside does not offer regular or sweet espresso drinks.

TABLE 6

Selected Own- and Cross-Price Elasticities within Shops

Price increase: Higher Grounds 12 oz. Drip Coffee	
Substitute good (at Higher Grounds)	
8 oz. Drip Coffee	0.468
12 oz. Drip Coffee	-4.345
16 oz. Drip Coffee	0.839
12 oz. Regular Espresso	0.019
12 oz. Sweet Espresso	0.019
Price increase: Higher Grounds 12 oz. Regular Espresso	
Substitute good (at Higher Grounds)	
8 oz. Regular Espresso	1.351
12 oz. Regular Espresso	-5.682
16 oz. Regular Espresso	1.482
12 oz. Drip Coffee	0.008
12 oz. Sweet Espresso	0.001
Price increase: Starbucks 12 oz. Drip Coffee	
Substitute good (at Starbucks)	
8 oz. Drip Coffee	0.192
12 oz. Drip Coffee	-5.210
16 oz. Drip Coffee	0.320
20 oz. Drip Coffee	0.353
12 oz. Regular Espresso	0.001
12 oz. Sweet Espresso	0.004

TABLE 7

Alternative Specifications

Alternative models		Original Model	Correlated β_{ir} and β_{is}	$\beta_{xi} = \beta_x v_i$ $v_i \sim \chi^2$	$\frac{N}{2}$	Partial Data
Specification		A	B	C	D	E
Parameter point estimates						
α : Price		-4.509	-4.561	-4.557	-4.049	-6.790
δ : Distance		-1.802	-1.701	-1.276	-1.687	-3.151
β_d : Drip coffee mean		0.388	0.426	0.501	0.514	1.772
β_r : Regular espresso mean		0.084	0.149	1.317	0.305	4.303
β_s : Sweet espresso mean		1.213	3.580	1.555	2.280	6.528
Price elasticity for Higher Grounds 12 oz. drip coffee		-4.345	-4.420	-4.714	-3.897	-5.980
Distortions at Higher Grounds						
Drip Coffee	8 oz.	1.222	1.233	1.483	1.304	3.804
	12 oz.	1.610	1.655	2.013	1.486	3.479
	16 oz.	1.888	1.966	2.509	1.599	3.208
Regular Espresso	8 oz.	5.972	6.114	5.671	5.481	7.051
	12 oz.	5.384	5.487	5.623	5.144	5.537
	16 oz.	3.901	3.953	4.294	3.713	3.449
Sweet Espresso	8 oz.	3.952	3.058	3.767	3.326	2.818
	12 oz.	2.756	2.297	2.783	2.367	2.893
	16 oz.	0.534	0.428	0.711	0.352	1.002

TABLE A1

Cumulative Class Enrollment

	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
Morning (8am – 2pm)	18780	17482	20056	17579	12318	143	0
Afternoon (2pm – 6pm)	7928	8479	8540	7658	2134	0	0
Evening (6pm – 12am)	1788	1272	1357	1050	1	0	0

TABLE A2

Labor Cost Regressions
Dependent Variable: $wage_k \times hours_t$

Specification	A	B	C	D
All transactions			0.162 (0.139)	
Drip coffee transactions	0.149[1] (0.047)			0.386 (0.390)
Espresso drink transactions	0.381[1] (0.070)			-0.356 (0.664)
Ounces of drip coffee (tens)		0.160[1] (0.038)	0.003 (0.150)	-0.164 (0.341)
Ounces of espresso drinks (tens)		0.256[1] (0.049)	0.151 (0.098)	0.513 (0.504)
R^2	0.920	0.919	0.920	0.920
F-statistic	452.06	387.07	480.02	404.64
Observations	71	71	71	71

Firm-specific fixed effects are included in each regression but omitted for confidentiality reasons.
1: Significant at the 99% level.

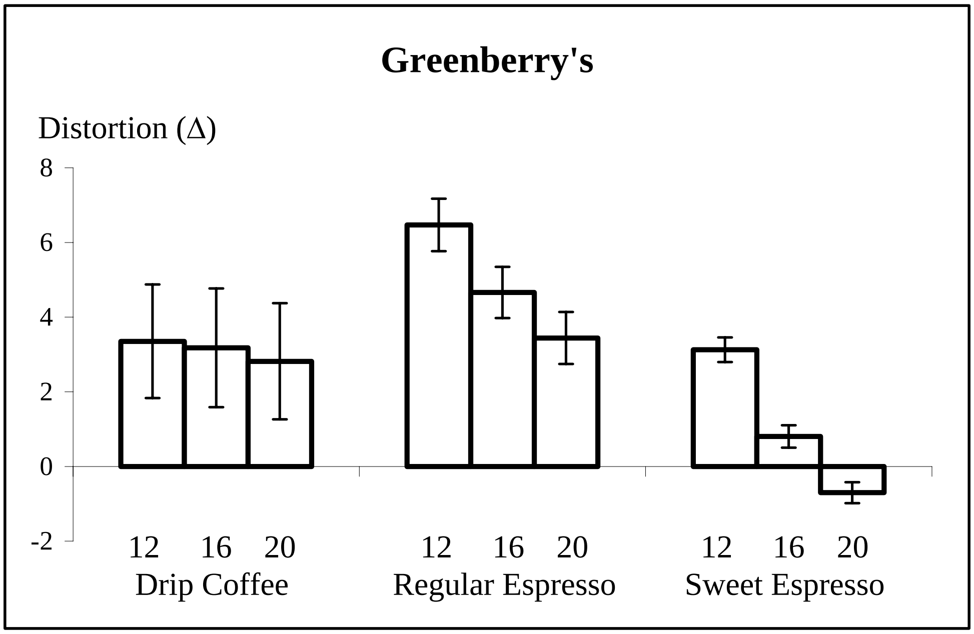
FIGURE 1

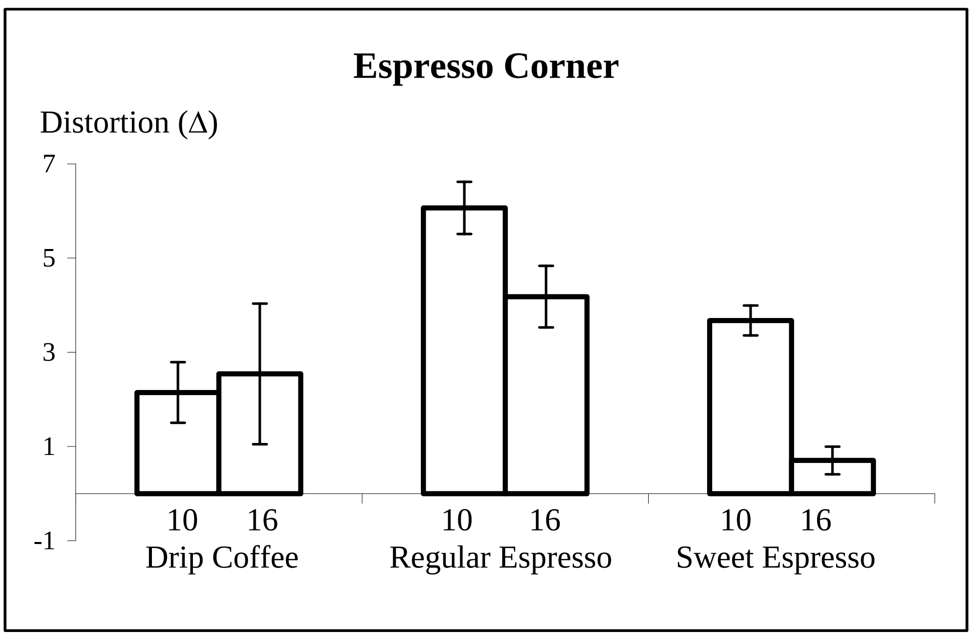
The Market for Specialty Coffee at the University of Virginia

The placement of the numbers in boxes corresponds to the locations of the following coffee shops. Greenberry's shops: Alderman [1], Bookstore [2], Poolside [3], and Tuttle [4]. Shops on the UVa 'Corner': Starbucks [5], Espresso Royale Caffe [6], and Espresso Corner [7]. Higher Grounds shops: East Cafeteria [8], and West Cafeteria [9]. The other lines on the map are outlines of buildings. The UVa Rotunda, the main landmark on campus, is between [1] and [9].

FIGURE 2

Distortions in Product Size

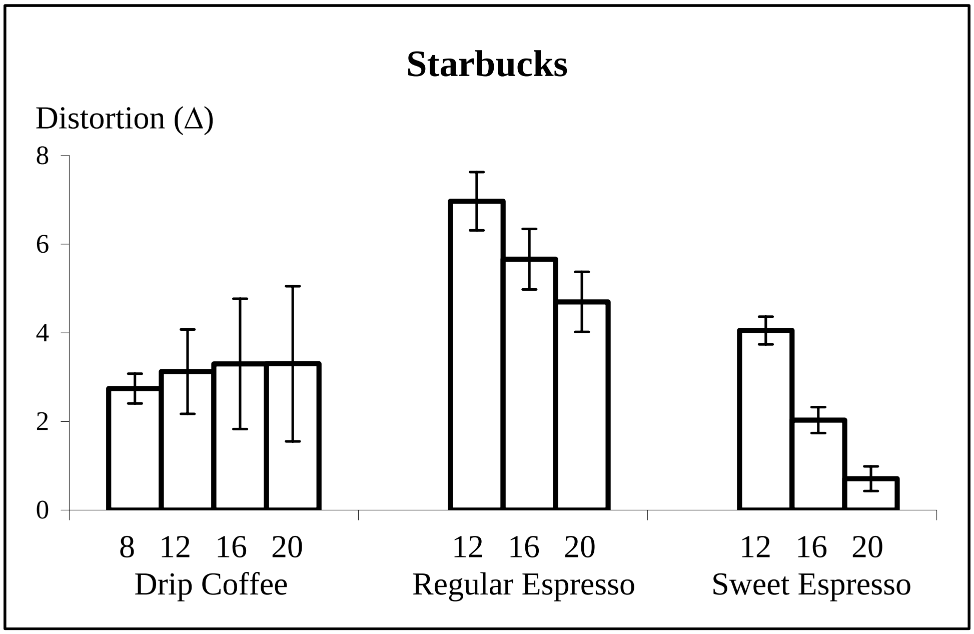


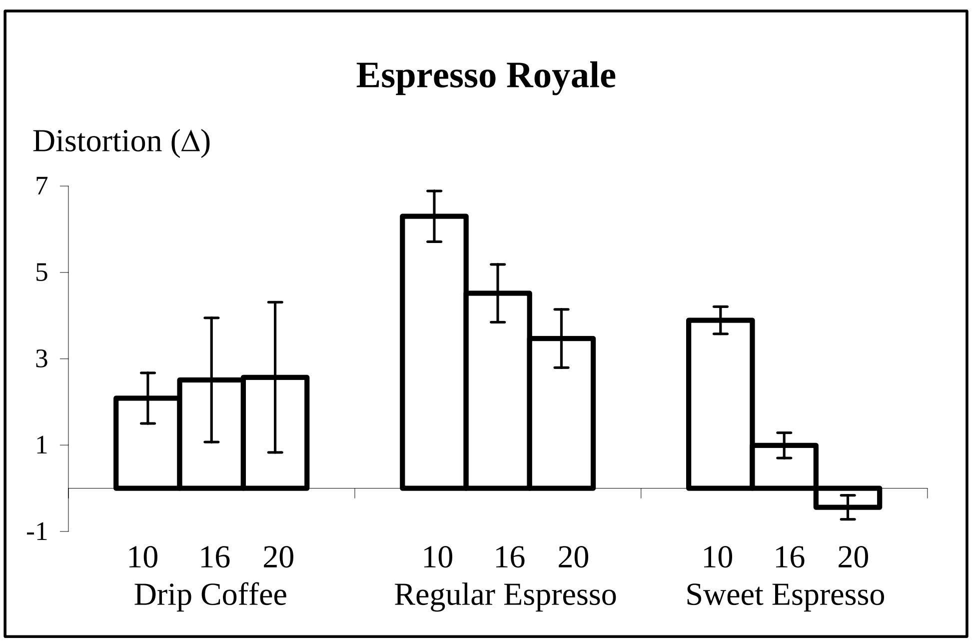


Notes for Figure 2:
Vertical axes measure Δ_j in pennies. 95% confidence intervals illustrated with error bars.

FIGURE 2
(continued)

Distortions in Product Size

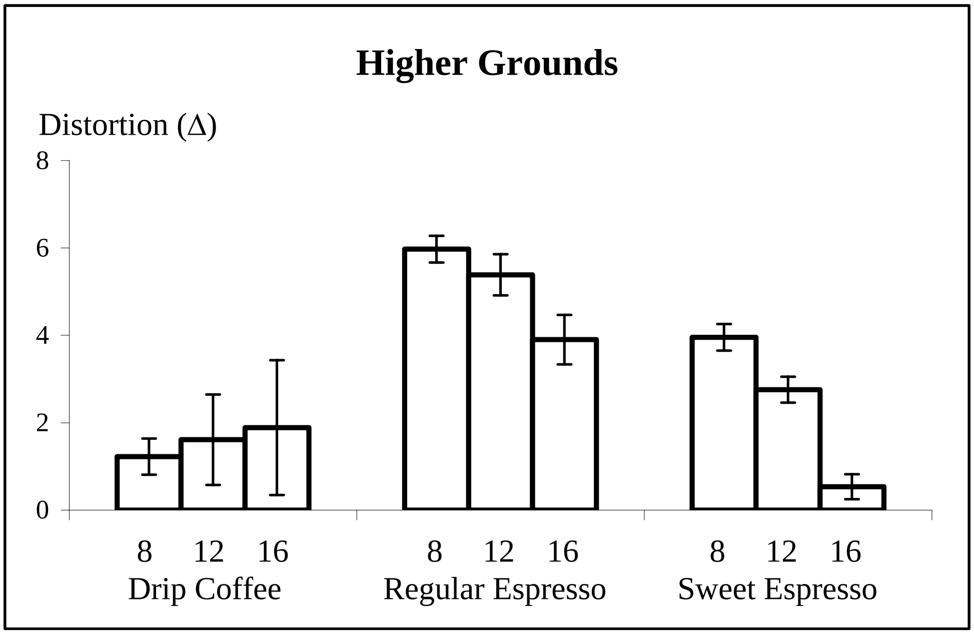




Notes for Figure 2:
Vertical axes measure Δ_j in pennies. 95% confidence intervals illustrated with error bars.

FIGURE 2
(continued)

Distortions in Product Size



Notes for Figure 2:
Vertical axes measure Δ_j in pennies. 95% confidence intervals illustrated with error bars.