Bundle-Size Pricing as an Approximation to Mixed Bundling

By Chenghuan Sean Chu, Phillip Leslie, and Alan Sorensen*

Multiproduct firms can set separate prices for all possible bundled combinations of its products ("mixed bundling"). However, this is impractical for firms with more than a few products, because the number of prices increases exponentially with the number of products. We find that simple pricing strategies are often nearly optimal. Specifically, we show that bundle-size pricing—*setting prices that depend only on the size of bundle purchased—tends to be more profitable than offering the individual products priced separately and tends to closely approximate the profits from mixed bundling.* (JEL D24, D42, L11, L13, L25)

We study the pricing problem of a multiproduct firm facing consumers who may purchase more than one (and possibly all) of the firm's products. Examples include cable television companies, professional sports teams, and online music stores. Such firms can choose from a wide variety of alternative pricing schemes. They can simply sell their products at a uniform price, or they can set different prices for each of their products. There are also bundling possibilities: the products could be offered only as a complete bundle, or subsets of products could be offered as bundles, and other products could be sold individually. The sheer number of available alternatives (for a firm with K products, there are $2^K - 1$ possible combinations of products that can be separately priced) makes this a highly complex problem for firms. Even for a firm selling only ten products, there are over 1,000 prices that could potentially be set.

In reality firms almost never implement complex pricing structures. Indeed, the reverse seems more common: firms often employ remarkably few prices. Why is this? In this study we show that simple pricing strategies are often nearly optimal. That is, in a broad class of models it takes surprisingly few prices to obtain nearly as much profit as would be earned by pricing every possible bundle combination. Of course, it matters *which* prices the firm chooses to set. We find that *bundle-size pricing*—a simple strategy that has not yet been explored in the literature—tends to be more profitable than offering the individual products priced separately, and tends to very closely approximate the profits from mixed bundling.

Bundle-size pricing (BSP) involves setting different prices for different sized bundles. For a firm with three goods, BSP sets one price for the purchase of any single good, a second price for the purchase of any two goods, and a third price for purchasing all three. The prior literature on bundling has ignored BSP, instead focusing on a few other alternatives: mixed bundling (MB), in which the firm chooses prices for every combination of goods; component pricing (CP), in which the firm sets different prices for each of its products; and pure bundling (PB), in which consumers' only option is to purchase all of the firm's products at a single price.[1] Our analysis focuses primarily on BSP and CP as competing approximations to MB.

The prior research offers two results of relevance for a firm with K products. First, MB tends to be strictly more profitable than CP.[2] Second, it is possible that PB is more profitable than CP.[3] Hence, the implication for a firm with ten products, say, would be: the best thing to do is set 1,023 prices under MB; and if that is not feasible (likely) then offering all products only as a single package may be more profitable than offering them individually (or perhaps not). Our findings offer a new suggestion: BSP requires only ten prices (the same number as under CP) and attains almost the same profit as MB under most circumstances—even when demand is highly asymmetric across products.[4] This is a significant step forward in providing practical advice for multiproduct firms.

We show that BSP and MB both tend to drive consumers to purchase larger-sized bundles than they would under CP. This has the effect of reducing consumers' heterogeneity in valuations for the products, which was always the key insight of the bundling literature. Put differently, the demand for each of the firm's K products under BSP (where a product is defined by bundle size) tends to be less heterogeneous than the demands for the K products under CP. With less heterogeneity, the firm can extract more surplus.[5] It may seem that CP would be more profitable when there is a high degree of demand asymmetry across products. Indeed, in the extreme case where each consumer demands only one of the firm's products (and average willingness to pay varies across products), bundling-based strategies are obviously less profitable than CP.[6] However, even in this situation, BSP does not necessarily do much worse than CP because it can extract surplus from individuals with high demand for only one product by setting a high price for single-good bundles.

The heterogeneity-reduction effect of bundling also implies that different bundles of the same size do not need to be priced very differently if the bundles are large. Hence, prices for large-sized bundles under BSP tend to be very close to prices under MB. This is why BSP tends to be a good approximation to MB. One interpretation of our findings, then, is that many of the prices a firm would set under MB are redundant.

---

[1] Component pricing is sometimes also referred to as "independent good pricing."

[2] See R. Preston McAfee, John McMillan, and Michael D. Whinston (1989).

[3] See George J. Stigler (1963) and William J. Adams and Janet L. Yellen (1976).

[4] Since BSP nests PB, it is also the case that BSP is always at least as profitable as PB and is often significantly more profitable.

[5] In the extreme, if marginal costs are zero, as $K \to \infty$, MB pushes all consumers toward purchasing the full bundle (i.e., MB simply implements PB in this limit) and extracts the entire consumer surplus. See Yannis Bakos and Erik Brynjolffson (1999).

[6] For example, a firm selling many different qualities of the same good would not want to employ a bundling strategy like BSP.

Our analysis has two components. First, we perform a large number of numerical experiments covering a broad range of demand and cost scenarios. In each experiment we compute the optimal prices under CP, PB, BSP and MB, and the associated profits. From a practical standpoint, numerical analysis is necessary for this problem because the profit maximization problem is analytically intractable under all but the simplest assumptions about the distribution of consumers' tastes.[7] Numerical analysis also allows us to explore a wider range of distributional assumptions and, thus, characterize underlying characteristics of demand and costs that enhance the profitability of BSP relative to CP in a way that is robust to functional-form assumptions.

The numerical analysis reveals that increasing the number of goods or decreasing marginal costs tends to make BSP more profitable than CP. Negative correlation in tastes also favors BSP, but only when the variance in consumers' tastes is roughly symmetric across goods and consumers' tastes are well above marginal costs. Otherwise negative correlation favors CP. We also show that increasing demand asymmetry has a nonmonotone impact on the relative profits of BSP and CP. These findings are counter to the conventional view that positive correlation and demand asymmetry *always* reduce the profitability of bundling relative to CP, and we devote significant analysis and discussion to explaining these exceptions.

A limitation of the numerical experiments is that we cannot be certain our results will transcend the particular parameter values we covered. For this reason, the second component of our analysis utilizes an estimated model, allowing us to demonstrate that our findings apply in an empirically relevant setting. The empirical analysis is based on a theater company that produces a season of eight plays. It is an interesting setting in which to compare the profitability of different pricing schemes because the plays differ in overall popularity, suggesting that component pricing may be important for profits, but many consumers attend multiple plays, suggesting that some form of bundling may also be profitable. With eight goods, MB would require the firm to set 255 prices, which is clearly impractical. In considering simpler alternatives, how important is it for the firm to set high prices for high-demand plays? What about offering discounts to consumers that attend multiple plays? Or some combination of these? And how do these alternatives compare with MB in terms of profits and consumer surplus?

A key feature of the theater data is that we observe the set of plays chosen by each customer. This allows us to identify the covariance in the joint distribution of consumers' tastes, which is an important determinant of profitability under alternative bundling schemes. The estimated demand system reveals strong positive correlations in tastes for most pairs of plays, which tends to reduce the relative profitability of bundling-type strategies compared to CP. Indeed, PB is 6 percent less profitable than CP in this case. Nevertheless, we find that BSP is 0.9 percent more profitable than CP, and BSP attains 98.5 percent of the MB profits.

While our focus is on bundling, approximating complex strategies using simpler alternatives is also an important issue in the closely related theory of contracts. William P. Rogerson (2003) argues that in a standard principal-agent model, most of the gains to the principal from offering the optimal continuous menu of contracts

can be captured by simpler alternatives.[8] McAfee (2002) and Robert Wilson (1993) establish similar findings in the context of nonlinear pricing, as does Zvika Neeman (2003) in the context of auctions.

I. Prior Theoretical Literature

The bundling literature explores the idea that a multiproduct monopolist can increase profits by selling goods in bundles, even when there are no demand-side complementarities or supply-side economies of scope. If a firm sells two products, and consumers vary in their willingness-to-pay for each product, then Stigler (1963) shows by example that selling these two products as a bundle (PB) may yield higher profit than if sold separately (CP). Adams and Yellen (1976) introduce MB as an alternative to CP and PB, showing by example that MB can strictly dominate both CP and PB. They also explain why higher values of marginal cost tend to favor CP over PB: with bundling, individuals may consume products for which their willing-ness-to-pay is less than the marginal cost to the firm.[9]

Two subsequent papers show that bundling (PB or MB) dominates CP in a wide variety of circumstances. First, Schmalensee (1984) expands the analysis to demand systems where consumers' product valuations are drawn from a bivariate normal.[10] Due to the limited computer power at the time, Schmalensee does not compute optimal MB prices, instead focusing on CP and PB. His main finding is that PB can be more profitable than CP, even when the correlation of consumers' valua-tions is nonnegative.[11] Second, McAfee, McMillan, and Whinston (1989) extend the prior results by showing that MB strictly dominates CP under rather general circumstances.[12]

All the above papers analyze two-product monopoly problems. A few prior papers study bundling with more than two goods. Bakos and Brynjolfsson (1999) focus on the profitability of PB as the number of goods (K) goes to infinity. They show that if goods have zero marginal cost, and if tastes are i.i.d. across products, then as K goes to infinity PB approximates perfect price discrimination.[13] This finding is particularly interesting in our context, since it provides an example of an incomplex alternative to MB that closely approximates the profitability of MB in a particular circumstance (i.e., large K).

---

[8] In the specific case of uniform types and quadratic effort costs, he shows that exactly 75 percent of the gains from offering the optimal menu can be obtained using a simple two-item menu, consisting of a fixed-price contract and a cost-reimbursement contract.

[9] In the language of Adams and Yellen (1976), these are violations of the *exclusion* condition.

[10] A concern with this approach is that the bivariate normal implies negative valuations for some consumers which would impact the analysis in nontrivial ways, as noted by Salinger (1995). In all of the analysis in our study we allow for free disposal, which has important consequences for bundling schemes, as we discuss in detail below.

[11] The numerical examples in Stigler (1963) and Adams and Yellen (1976) somehow suggest the importance of negative correlation, as noted by Schmalensee (1984).

[12] McAfee, McMillan, and Whinston (1989) also distinguish between firms that can monitor purchases or not. With monitoring, the firm can charge a price for the bundle of two that is higher than the sum of component prices. We limit our analysis to the no-monitoring case. See also Alejandro M. Manelli and Daniel R. Vincent (2006), who point out circumstances under which even MB can be dominated by pricing mechanisms involving random assignments.

[13] Bakos and Brynjolfsson (1999) also show that, under certain conditions, increasing the number of goods under PB monotonically increases profit. Xianjun Geng, Maxwell B. Stinchcombe, and Andrew B. Whinston (2005) extend the analysis of Bakos and Brynjolfsson to incorporate diminishing marginal utility.

Mark Armstrong (1999) provides a more general but similar result to that of Bakos and Brynjolfsson (1999). He shows that a two-part tariff, in which consumers are charged a fixed fee and can then purchase any products at marginal cost, achieves approximately the same profit as perfect price discrimination if the number of products approaches infinity. In the special case of zero marginal cost the two-part tariff is equivalent to PB. The focus on settings with large numbers of products may be quite relevant for some firms, such as booksellers or supermarkets. But clearly these results are of questionable relevance to firms with five products, for instance.

Fang and Norman (2006) also examine the profitability of PB with more than two goods. In contrast to Armstrong (1999) and Bakos and Brynjolfsson (1999), they focus on finite K, and they seek to determine under what circumstances PB is an attractive pricing strategy. They confirm that increasing marginal cost tends to favor CP over PB, as Adams and Yellen (1976) had argued. They also show (by way of numerical experiments) that increasing the number of goods may favor PB over uniform pricing (UP).[14]

For a firm selling a finite number of goods, the prior literature can be easily summarized: MB is *always* more profitable than CP, and in *some* cases PB may also be more profitable than CP. We contend these results are of limited practical value—MB rapidly becomes impractical as the number of goods increases above a mere few, and even in the cases when PB is more profitable than CP, it is conceivable there are other straightforward pricing schemes that will do even better.

Hence, we focus on the question: do pricing schemes exist that involve few enough prices to be feasible, and that tend to yield profits close to the MB level?

II. The Multiproduct Pricing Problem

In principle, multiproduct firms can choose from a wide variety of pricing schemes. For a firm with K products, the optimal MB strategy requires setting $2^K - 1$ prices.[15] PB and UP require only one price to be set: the price for the bundle of all K products (in the PB case), or the per-product price (in the UP case). In between these extremes are CP, by which we mean setting K different prices for the K different products, and BSP, by which we mean setting K prices that depend on the number of products purchased. Note that MB nests all the simpler pricing strategies as special cases, so it will always be weakly more profitable than any of these alternatives. Similarly, CP nests UP as a special case, and BSP nests both UP and PB as special cases. CP and BSP are nonnested alternatives.

It is natural to ask: what is the most profitable pricing plan for a given number of prices? In other words, it would be interesting to compute the upper bound on profit from any pricing strategy involving N prices, for each value of N from 1 to $2^K - 1$, given a particular model of demand and costs.[16] Such an upper bound is

---

[14] UP is defined as each product sold separately at the same unit price.

[15] We subtract 1 because the firm does not set the price for the outside good.

[16] David A. Malueg and Christopher M. Snyder (2006) show a related result: if a monopolist sells to N independent markets with different demands, and the cost function is superadditive (plus some other assumptions), then the ratio of CP profit to UP profit is at most N. Also, Armstrong (1999) quantifies the degree to which increasing the number of the goods (K) allows the optimal two-part tariff scheme to capture a larger fraction of the maximum possible profit.

obviously increasing in N, but it would be useful to know how rapidly the upper bound approaches the MB level of profits. This would provide an indication of the value to the firm from additional complexity (as measured by the number of prices).

There are a couple of challenges to computing such an upper bound when goods are discrete and tastes for products can be arbitrarily distributed. First, there is a conceptual issue: what does it mean to say a firm can choose any N prices? One needs to be precise about how to determine the implied prices of bundles with unspecified prices.[17] Second, in any example with more than a few goods, a large number of possible pricing structures must be evaluated, which is a nontrivial computational challenge. For these reasons, a complete answer to this question is beyond the scope of this study.

Some suggestive implications come from Wilson (1993), who discusses nonlinear pricing for products with multiple continuous characteristics. Wilson establishes necessary conditions and a numerical method for computing the optimal monopoly solution and finds that pricing based on some measure of the aggregate "size" of the purchase is a common feature.[18] In the special case when there is a single product characteristic, the loss in profits from using an N-part tariff (relative to the fully optimal nonlinear tariff) is on the order of $1/N^2$. For what we are interested in—i.e., discrete products—the analog to having a single product characteristic is for tastes for different products to be perfectly correlated. Wilson's proof relies on differentiability and thus cannot be applied to a discrete combinatoric problem. But we might conjecture that here, too, a few prices are all that is necessary to closely approximate mixed bundling profits when tastes for different bundled goods are strongly correlated.[19]

It should be noted that in our numerical experiments BSP usually attains around 99 percent of the profit from MB. Hence, it is already clear from our analysis that the upper bound on profits does indeed rapidly increase in the number of prices, N. There may exist other pricing schemes involving K prices that are more profitable than BSP in any particular example, but typically such schemes can yield at most a 1 percent improvement.

In practice, multiproduct firms tend to use a broad range of different pricing/bundling strategies. Consider baseball teams, for example, which have 81 home games (products).[20] For the 2006 season, the Los Angeles Dodgers offered several bundles of specific games, a discount for choosing any 27 games, and equal prices for all individual games. In contrast, for 2006 the San Francisco Giants did not offer any bundles or quantity discounts, but did vary prices by day of week and by opponent. Variation in pricing strategies is also evident in settings with fewer products. Consider the Steppenwolf Theater in Chicago that produces a five-play season. In 2006–2007 they offered a discount for the five-play bundle at a variety of prices that varied by time of week, and equal prices for individual shows (also varying by time

---

[17] For example, CP involves K prices for each individual good *and* a rule for constructing the price of any bundle with two or more goods. Namely, the price of the bundle is the sum of the included components' prices.

[18] See Chapter 13 for the multidimensional case and Chapter 6 for the one-dimensional case.

[19] When tastes are perfectly correlated, BSP can in fact be thought of as being equivalent to a K-part tariff. It is also easy to prove that in this case CP can replicate the optimal K-part tariff and thus does at least as well.

[20] Under MB, with 81 products a firm would set 2.4×10^{24} prices.

of week). In 2006–2007 the San Francisco Opera had a ten-opera season and offered 37 bundles (combinations of specific operas and time of week), and equal prices for individual shows (also varying by time of week). These examples highlight the dramatic differences in pricing strategies implemented by different firms in similar settings. We have been unable to find an example of MB being used in practice for three or more products.

A. *An Example with Two Goods*

In order to clarify the differences between BSP, CP, and MB, in this section we present a straightforward example with analytic solutions. Here, and throughout the remainder of the paper, we assume the firm is a monopolist and adhere to the standard assumptions of the bundling literature:

   (i)   Consumers purchase one or zero units of each product;

   (ii) Consumers' valuations for a bundle equal the sum of their valuations for the bundle's component products (i.e., products are neither complements nor substitutes); and

   (iii) There is no resale.

There are two goods, both with zero marginal cost. Consumers' valuations for good 1 are uniformly distributed between 0 and θ: $v_1 \sim U[0,\theta]$. And consumers' valuations for good 2 are uniformly distributed between zero and 1: $v_2 \sim U[0, 1]$. Also assume that v_1 and v_2 are uncorrelated. The virtue of this model is that we can derive analytic solutions for the optimal prices under CP, BSP, and MB, as well as the associated profits (see Appendix B for details).[21] If $\theta = 1.7$ then the optimal CP prices are 0.85 and 0.5 for goods 1 and 2, respectively. Under BSP, the optimal price for a single-good bundle is 0.9, and the price for the bundle of both goods is 1.1. In this example BSP is 5.6 percent more profitable than CP (even though the optimal CP prices vary by 70 percent across the two goods).

The comparison between BSP and MB in this example is also instructive. Under MB, the price for good 1 is 1.13, the price for good 2 is 0.67, and the price for the bundle is 1.18. Unsurprisingly, the price for a single-good bundle under BSP (0.9) lies between the two single-good prices under MB. The price for the two-good bundle is quite close under BSP and MB (1.1 and 1.18 respectively), in comparison to the differences in single-good prices. However, only 15 percent of the total profit under MB comes from consumers who buy one good, with the remaining 85 percent coming from sales of the bundle. This pattern of BSP prices closely approximating the MB prices for bundles but not for individual goods, and of bundles contributing more to profits, also applies to the numerical experiments we analyze below. Put simply, BSP prices closely approximate MB prices where it matters most—for large-sized bundles.

---

[21] A limitation of this model is that BSP is weakly more profitable than CP for all values of θ. Nevertheless, the model is helpful for demonstrating the differences between CP and BSP for a given value of θ.
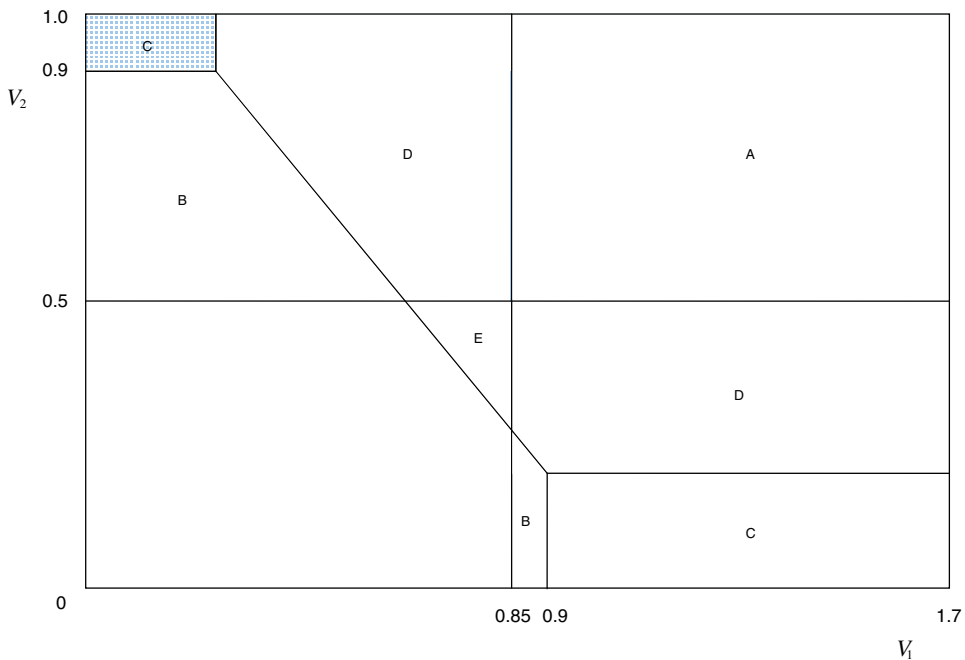
FIGURE 1. SEPARATION OF CONSUMERS UNDER CP AND BSP

In Figure 1 we show how CP and BSP lead to different partitions of consumers (for $\theta = 1.7$). CP is the most straightforward: consumers to the right of 0.85 purchase good 1, and consumers above 0.5 purchase good 2 (with consumers in region A purchasing both). Under BSP, consumers in the two regions labeled C purchase one product (good 1 for the lower right C, and good 2 for the upper left C). And under BSP consumers in regions A, D, and E choose the bundle of both goods.

One interesting way to read Figure 1 is to ask which pricing scheme extracts more surplus from which consumers. Consumers located in region A purchase both goods under CP and BSP. Under CP, these consumers each pay 1.35 and under BSP they pay 1.1. Hence, the firm extracts more surplus from consumers in region A by using CP rather than BSP. CP also extracts more surplus from consumers in region B, since these consumers buy either good under CP and buy nothing under BSP. BSP, on the other hand, extracts greater surplus from consumers in regions C, D, and E. Region E is particularly interesting because these consumers purchase nothing under CP, and under BSP they purchase the bundle of both goods. Consumers in region D also increase the number of goods purchased (from a single good to two). In region C the number of goods consumed remains at one, but BSP extracts more surplus because the price for a single good (0.9) is greater than both single prices under CP. To summarize these differences, in Figure 1 we shade the regions in which BSP extracts more surplus from consumers than CP.

There are four main points to take from Figure 1. First, BSP is more focused than CP on getting consumers to purchase multiple goods. Relative to CP, BSP raises the

price for single-good buyers and lowers the price for multigood buyers. It is profitable to do so in this example, but there is a downside:

(i) By increasing the price for a single-good bundle, some consumers are excluded from purchasing anything who otherwise would have purchased something (region B); and

(ii) Consumers who would have purchased both goods under CP are given a discount under BSP with no change in their purchase choice (region A).

Second, from the figure it is apparent why negative correlation in consumers' valuations may increase the relative profitability of BSP. Note the downward trend of the shaded regions in which BSP extracts greater surplus than CP—negative correlation tends to increase the fraction of consumers in these regions.[22] It is also apparent from the figure that BSP is capable of extracting more surplus from individuals in the tails with high valuations for one product and low valuations for the other (region C). Hence, it is wrong to presume that BSP is poor at extracting surplus from consumers with a high valuation for only one product.

The third point concerns the consequences of diminishing marginal utility. The model that underlies Figure 1 assumes the utility of the bundle equals the sum of the utilities of the two goods. For some products, however, it may be important to incorporate diminishing marginal utility into the analysis—e.g., by lowering the utility of a bundle by some factor that increases with the size of the bundle. In the extreme, if diminishing marginal utility is so strong that individuals never consume more than one good, then CP is weakly more profitable than BSP (for any distribution of valuations).

Such reasoning suggests that any degree of diminishing marginal utility should reduce the profitability of BSP relative to CP, since the value of bundles is lowered. But this is wrong. Diminishing marginal utility also reduces the profitability of CP, possibly by even more than it does for BSP. This is because CP also benefits from extracting surplus from individuals who purchase both goods (region A), and CP actually extracts more surplus from this set of consumers than BSP does. In other words, diminishing marginal utility reduces willingness-to-pay for the bundle of both goods, which also reduces the amount of surplus that CP can extract. In the counterfactuals based on our empirical analysis in Section IV, we indeed verify that incorporating diminishing marginal utility can reduce the profitability of CP by even more than it does BSP. This is another appealing aspect of BSP from a firm's point of view.

Fourth, this example gives an indication of the complexity of the BSP pricing problem. In this simple case with two goods and independent, uniformly distributed taste distributions, the regions of integration for determining demand for different-sized bundles are nonrectangular and noncontiguous (i.e., region C). Adding more

---

[22] Introducing correlation to the example will change the optimal BSP prices, changing the details of the figure. Note, however, the optimal CP prices do not depend on the correlation of consumer's valuations—each good is optimally priced independently of the other good, so correlation plays no role in the CP optimization problem. Hence, the figure would change in some ways, but it would be qualitatively similar, and this point would still hold.

TABLE 1—ALTERNATIVE PRICING STRATEGIES

Initials	Name	Num. prices	Description
UP	Uniform pricing	1	Each product sold separately at a uniform price
PB	Pure bundling	1	Only option for consumers is the full bundle
CP	Component pricing	K	Each product sold separately at a different price
BSP	Bundle-size pricing	K	Prices depend only on number of purchased products
MB	Mixed bundling	$2^K - 1$	Separate prices for every possible combination of products

goods or incorporating nonzero correlation will increase the complexity, and allowing for more realistic distributions of valuations (such as normal) precludes analytic solutions. This is why numerical methods are essential for solving the BSP optimization problem in more general settings.

B. *Numerical Analysis with Continuous Types and More Than Two Goods*

In this section we examine a broad range of computational experiments in which we solve for the optimal prices and profits for five different pricing strategies, which are detailed in Table 1. In all experiments we assume a demand model in which consumer i's utility from purchasing bundle j is equal to $V_i'D_j - p_j$, where V_i is a $K \times 1$ vector of valuations for the firm's K products, D_j is a $K \times 1$ vector of binary indicators for which of the K products are included in bundle j, and p_j is the price of bundle j.[23] Each consumer's problem is to choose the offered bundle that maximizes her utility. Consumers' product valuations are heterogeneous: V_i is drawn from a (multivariate) distribution F. We vary the number of goods in the experiments from two to five. Importantly, we allow for free disposal—if a consumer purchases a bundle that includes a product for which she has a negative valuation, we assume zero utility from consuming that product.[24]

We also incorporate cost variation into the experiments. This is important because, as noted by Adams and Yellen (1976), higher values of marginal cost should favor CP over PB. Hence, it is conceivable that higher values of marginal cost should also favor CP over BSP. Each experiment is therefore performed under four different assumptions regarding costs:

  (i) All products have zero marginal cost;
 (ii) All products have positive and equal marginal cost (we set the marginal cost equal to 0.2);
(iii) All products have positive but differing marginal cost (we set marginal costs equal to half of the product's mean valuation); and
 (iv) Marginal costs are zero but there is a binding capacity constraint.[25]

---

[23] As in the above example, by assuming additive preferences we are ruling out consumption complementarities as a motivation for bundling.

[24] Schmalensee (1984) does not allow free disposal. Either assumption may be correct depending on the particular products. As we will explain, the free-disposal assumption has some striking consequences for the analysis.

[25] For the experiments with capacity constraints we first find the optimal uniform price in the absence of any capacity constraint, and then set the capacity constraint equal to 0.9 times the demand for the most popular product

TABLE 2—ALTERNATIVE TASTE DISTRIBUTIONS

Name	Description
Exponential	Marginal distributions of the v_{ik}s are exponential, with means between 0.2 and 2.0
Logit	Marginal distributions of the v_{ik}s are extreme value, with means between 0 and 2.5 and scale parameter $= 0.25$
Lognormal	v_k is a multivariate lognormal random variable; $log(v_{ik})$s have variance 0.25 and means between -1.5 and 1
Normal	v_i is a multivariate normal random vector; v_{ik}s have variance 0.25 and means between -1 and 2.5
Normal(v)	v_i is a multivariate normal random vector; v_{ik}s have mean zero and variances between 0.25 and 1.75
Uniform	Marginal distributions of the v_{ik}s are uniform on $[0, a_k]$, with a_k between 0.4 and 4

Notes: For each group of distributions we consider several combinations of parameter values, and for each set of parameter values we consider three covariance structures: negative correlation, independence, and positive correlation. For the normal distribution, we consider two additional covariance structures with a mix of positive and negative correlation. The details are described in online Appendix C.

Table 2 describes the alternative assumptions on the distribution of consumers' valuations (F) that we consider in our experiments. We consider five parametric families: exponential, logit, lognormal, and normal, which are all commonly used in empirical studies of demand; and the uniform distribution, which is often convenient in theoretical studies of demand and is also occasionally used in empirical work. For the normal distribution, we consider two cases: one in which we hold the variances fixed and vary the means, and another in which we hold the means fixed and vary the variances. For each family, we consider many sets of parameter values (as explained below), and for each set of parameter values, we consider three covariance structures: negatively correlated tastes, independent tastes, and positively correlated tastes. This is important since correlation in tastes is a key determinant of the profitability of bundling. For distributions where covariance is not readily parameterized (e.g., exponential, logit, and uniform), we use Gaussian copulas to make the consumers' valuations correlated across products while leaving the marginal distributions of valuations unchanged. The details are described in the Web Appendix (available online). Also, for the normal family only, we consider two forms of mixed correlation, in which tastes are negatively correlated for some pairs of products, and positively correlated for others.

For each parametric distribution we perform experiments for a broad range of parameter values. To help others reproduce our findings, rather than randomly draw parameter values, we define a grid of uniformly spaced parameter combinations. The grid boundaries for each parametric family are shown in Table 2. The boundaries were chosen so that the range of optimal prices is roughly similar across cases, to help with comparability, with optimal CP prices varying by up to a factor of 10.[26] Thus, our experiments include cases with a relatively high degree of demand asymmetry, which can favor CP (though, as we explain below, the effect is nonmonotone, and BSP and CP are equivalent in proportional terms when demand asymmetry is sufficiently high).

under the optimal uniform price. This ensures that the capacity constraint will be binding for at least one product under UP, regardless of the particular parameters of the taste distribution.

[26] Specifically, the range of parameters for each distributional family is such that the optimal component prices (assuming zero marginal cost) vary from about 0.2 to 2.0.

It is conceivable a firm may consider bundling together products for which the optimal component prices vary by much more than a factor of 10. We have chosen instead to focus on settings where the component products are more similar. Baseball games are perhaps an ideal example, because it is conceivable that the most popular regular-season game would have an optimal component price that is several times greater (though probably not more than ten times greater) than the least popular game. See also our empirical example in the next section. Nevertheless, some of our results may not generalize to settings where demand differs more dramatically across products.

For different numbers of products, we adjust the granularity of the grid of parameter values such that we end up analyzing approximately 220 parameter combinations for each class of distribution regardless of the number of products. The combination of six parametric families, three covariance structures for each (five for the Normal), four marginal cost assumptions, variation in the number of products from two to five, and about 220 parameter combinations in each case, leads us to compute five sets of prices and profits in 71,360 different examples. Numerical methods are used to find the optimal prices in each case.[27]

It is important to acknowledge the limitations inherent to this kind of computational analysis. Although we attempt to cover a large space of parameter values, the results clearly depend on the specific parameters we choose (i.e., the choice of grid). Further, there is no way for us to know whether we are under- or oversampling the relevant (i.e., empirically plausible) combinations of parameters. Nevertheless, by comparing across experiments, we can draw conclusions about the primitives of demand and costs that favor each pricing scheme.

In the next several subsections we describe and analyze the results from these numerical experiments. A thorough description of the results would be too lengthy for inclusion in this article. Instead, we summarize the most important features of the results here and provide a more thorough description—including a detailed series of tables—in the Web Appendix (available online). Just as importantly, we also seek to explain what primitives of demand and costs favor each pricing scheme.

Number of Goods.—Figure 2 summarizes *all* the numerical experiments and demonstrates the impact of changing the number of goods (K). The figure shows box plots depicting various percentiles of the distribution of profits under each pricing strategy, relative to BSP. To construct a given box plot we pool experiments across distributions of consumers' valuations and for different cost structures. Note that we pool across parameter combinations for a given parametric family as well as pooling across parametric families (in addition to pooling across cost scenarios). Hence, while each box plot reveals the range of outcomes, it hides differences across

---

[27] Specifically, we use SNOPT, a sequential quadratic programming algorithm developed by Philip E. Gill, Walter Murray, and Michael A. Saunders (2002) for solving nonlinear constrained optimization problems. Under BSP, the per-product prices are constrained to be nonincreasing in the size of the bundle. For BSP and MB, we also check to make sure the computed optimal prices are robust to alternative start values. We calculate the demands for each bundle using a kernel-smoothed frequency simulator, as discussed in Vassilis Hajivassiliou, Daniel L. McFadden, and Paul Ruud (1996), using 10,000 simulated consumers and a logistic kernel with smoothing parameter 0.02.
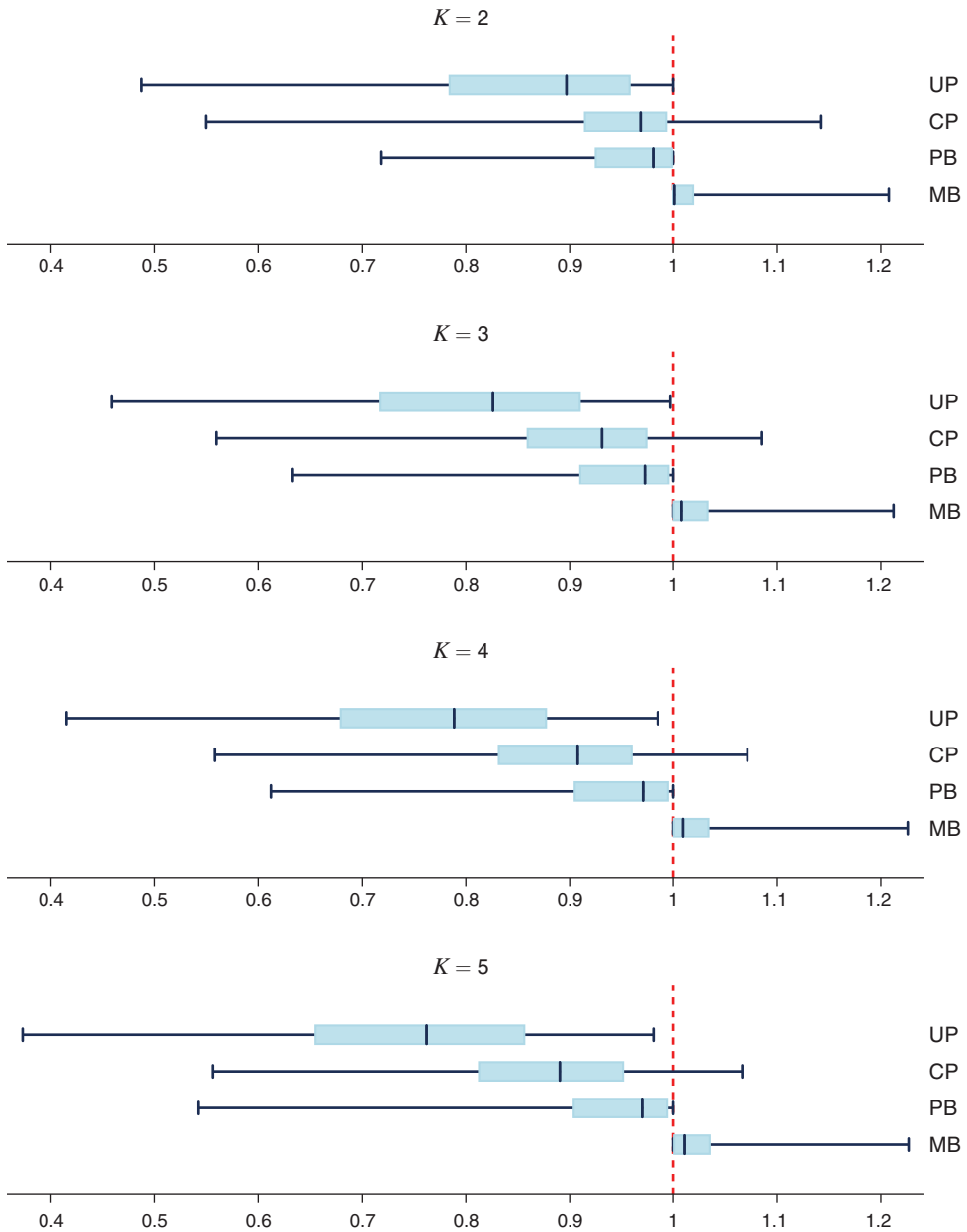
FIGURE 2. DISTRIBUTIONS OF PROFITS FOR EACH PRICING STRATEGY, RELATIVE TO BSP, FOR DIFFERENT VALUES OF *K*

Note: Each box plot depicts the 1st, 25th, 50th, 75th, and 99th percentile of the distribution of profit relative to the profit from BSP.

distributions and for varying costs (which we disentangle, below). In Figure 2 each box plot indicates the 1st, 25th, 50th, 75th, and 99th percentiles of the distribution of profit for a given pricing strategy relative to BSP.[28]

---

[28] We depict the 1st and 99th percentiles instead of the min and max of the distribution because occasionally optimization error leads to misleading values for these extremes.

As expected, Figure 2 shows that MB is always more profitable than BSP (because MB nests BSP), and BSP is always more profitable than UP (because BSP nests UP). However, there are two more substantive findings that are immediately apparent in Figure 2:

 (i) BSP tends to be more profitable than CP. Based on the 71,360 experiments we performed (across different cost assumptions and across different taste distributions), we find that BSP is more profitable than CP 90 percent of the time. Furthermore, BSP obtains 14 percent higher profit than CP, on average.

 (ii) BSP tends to obtain profits that are within 1 percent of the profits from MB. Specifically, the profit from BSP is within 1 percent of MB in 55 percent of the 71,360 experiments we performed. And on average, we find that BSP yields 98 percent of the MB profits.

To understand these findings, recall the bundling literature's basic insight that PB is sometimes more profitable than CP. Here we find that PB attains higher profit than CP in 62 percent of our numerical experiments. We also find that increasing the number of goods tends to favor PB over CP: for $K = 2,3,4,5$, PB is more profitable in 54, 62, 65, and 67 percent of the experiments, respectively. Fang and Norman (2006) also find this pattern in their numerical experiments. Since BSP nests PB, we expect BSP also becomes more profitable as K increases, relative to CP. Indeed, for $K = 2,3,4,5$, BSP is more profitable than CP in 81, 90, 93, and 95 percent of the experiments, respectively.

The reason why increasing K favors bundling-type schemes is the heterogeneity-reduction effect: as the number of goods in a bundle increases, the variance in consumers' valuations for the bundle tends to decrease. Hence, under MB the optimal prices for different bundles of a given size K are not very different from each other, provided K is large. Moreover, the average price under MB of size-K bundles is close to the uniform price for such bundles under BSP. This is why BSP prices tend to be an especially good approximation of MB prices for large-sized bundles.

This point is illustrated in Table 3, which compares CP and BSP as competing approximations to MB. The table shows how close the CP prices and BSP prices are to the MB prices, as well as how close the market shares are. For each possible bundle of a given size we compute price differences (as a percentage of the MB price), and average these differences across experiments. For example, based on all of our experiments with $K = 3$, including all cost scenarios, CP prices for individual component sales (bundle size equals one) tend to differ from the MB prices by 34.6 percent. In contrast, BSP prices in the same experiments tend to differ from MB prices by 52.7 percent. Hence, Table 3 reveals that CP prices for small-sized bundles (for any given K) tend to be closer than BSP prices are to the fully optimal MB prices. However, the table also shows that for large-sized bundles, and especially for the bundle of all K products, the BSP prices are typically very close to the MB prices, unlike CP.

Consider also the market shares shown in Table 3. Under BSP and MB the tendency is for the majority of consumers to purchase the bundle of all K products, while under CP there are relatively few sales of the full bundle. For example, with

TABLE 3—PRICE DIFFERENCES AND MARKET SHARES, BY BUNDLE SIZE

Bundle size	Average price differences		Market shares		
	$|p_{CP} - p_{MB}|/p_{MB}$	$|p_{BSP} - p_{MB}|/p_{MB}$	CP	BSP	MB
Panel A. K = 3					
1	0.346	0.527	0.366	0.094	0.118
2	0.172	0.151	0.258	0.163	0.192
3	0.176	0.038	0.110	0.345	0.333
Panel B. K = 4					
1	0.410	0.674	0.294	0.060	0.088
2	0.242	0.242	0.261	0.110	0.133
3	0.167	0.095	0.172	0.154	0.171
4	0.180	0.036	0.072	0.311	0.294
Panel C. K = 5					
1	0.457	0.802	0.237	0.043	0.067
2	0.300	0.328	0.239	0.080	0.098
3	0.212	0.154	0.193	0.104	0.120
4	0.171	0.079	0.120	0.143	0.160
5	0.190	0.038	0.051	0.284	0.265

Notes: Price differences are calculated as a percent of the MB price and then averaged across prices within bundle size and across experiments. Market shares are averages across experiments for bundles of a given size. For example, on average across experiments with $K = 3$, MB pricing leads 11.5 percent of consumers to purchase a single product.

$K = 3$, CP sells a single good to 37 percent of consumers, while BSP and MB sell a single good to only 9 percent and 12 percent, respectively. Meanwhile, BSP and MB sell the full bundle to 35 percent and 33 percent of consumers, respectively, and CP sells the full bundle to only 11 percent of consumers. Hence, pricing under CP tends to be a better approximation to MB for small-sized bundles than BSP, while BSP tends to be a better approximation to MB than CP for the large-sized bundles, but the large-size bundles matter more: MB tends to sell many more large-sized bundles than CP, and BSP does about as well as MB in this respect.

A potential concern is that BSP closely approximates MB merely because prices simply do not matter very much in our experiments. To examine this possibility, we computed MB profits in cases where the firm is mistaken about the distribution of consumers' valuations. Suppose the true distribution of consumers' valuations is joint normal with positive correlations, but the firm sets MB prices incorrectly assuming negative correlations of the same magnitude. In unreported experiments we found this tends to yield around 15 percent lower profit than if the firm had correctly assumed positive correlations.[29] This provides a degree of assurance that profits are indeed sensitive to prices in our experiments.

Correlation.—Figure 3 summarizes the results from the experiments separately for the three covariance structures we consider. Typically (i.e., at the medians), BSP performs best when tastes are negatively correlated. Interestingly, however, the figure also reveals that negative correlation increases the dispersion of outcomes for

---

[29] These experiments were performed for $K = 2,...,5$ and with both zero and positive marginal costs. We also considered a variety of other examples of mistaken beliefs. The results were qualitatively the same in all cases.
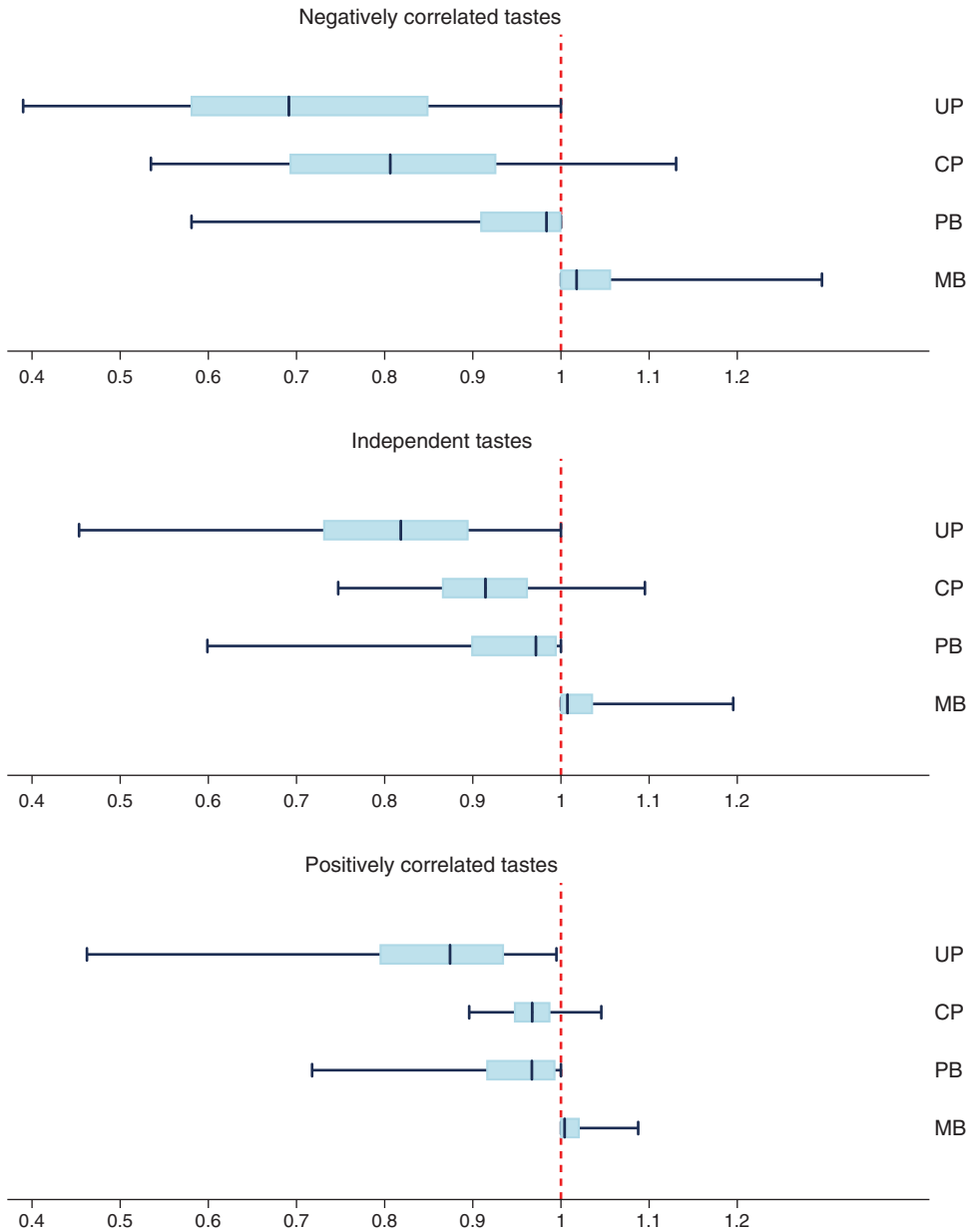
FIGURE 3. DISTRIBUTIONS OF PROFITS FOR EACH PRICING STRATEGY, RELATIVE TO BSP,
UNDER DIFFERENT COVARIANCE STRUCTURES

Note: Each box plot depicts the 1st, 25th, 50th, 75th, and 99th percentile of the distribution of profit relative to the profit from BSP.

relative profits. With negatively correlated tastes, CP can be up to 12 percent more profitable than BSP, and up to 47 percent less profitable. For independent or positively correlated tastes, the range of possibilities is much smaller, at *both* extremes. Surprisingly, in some cases negative correlation *reduces* the profitability of BSP relative to CP. Why is this?

FIGURE 4. BSP PROFITS RELATIVE TO CP PROFITS, AS A FUNCTION OF CORRELATION IN
CONSUMERS' TASTES ACROSS PRODUCTS

Notes: This figure plots the ratio of BSP profits to CP profits as a function of correlation. In each of the two cases shown, the taste distribution is bivariate normal with ($\mu_1, \mu_2, \sigma_1, \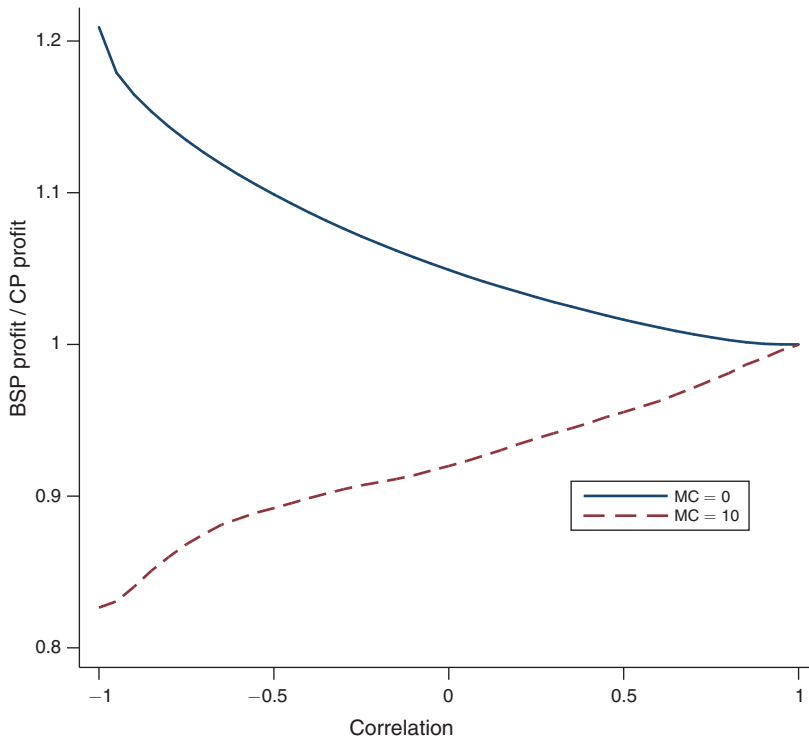sigma_2$) equal to (10,10,3,1). The difference between the two cases is that marginal costs are equal to zero for both products in one case and equal to 10 for both products in the other case.

The usual intuition is that negative correlation favors bundling because it accentuates the heterogeneity-reduction effect of bundling. This intuition is generally correct when the variance in consumers' tastes is roughly symmetric across goods and if consumers' tastes are well above marginal costs. However, to allow for situations in which BSP does poorly, we chose our parameters such that these two conditions often do not hold in our experiments. In such instances, we see that larger negative correlation may actually *reduce* the relative profitability of BSP. Specifically, if the taste distribution puts significant density on valuations *below* marginal cost, the relationship between BSP profits and correlation can reverse. A similar reversal also occurs if the variances for different goods are highly asymmetric.

This principle is illustrated in Figure 4, which plots the ratio of BSP profit to CP profit as a function of correlation for two levels of equal marginal costs for two products, holding fixed the marginal distributions of consumers' valuations.[30] Product valuations are normally distributed, with mean valuations equal to 10 and standard deviations equal to 3 and 1, respectively. When marginal costs are zero, the relative

---

[30] Note that it would be equivalent to plot BSP profits alone (instead of their ratio to CP profits). CP profits depend only on the marginal distributions of tastes and are thus unaffected by changes in the correlation.

profitability of BSP declines monotonically as a function of correlation in tastes: the more negatively correlated are consumers' tastes, the more profitable is BSP relative to CP. When marginal costs are 10, so that many consumers' product valuations are below marginal cost, the sign of the relationship completely reverses: as correlation becomes more negative the relative profitability of BSP to CP decreases.[31]

To understand why this reversal occurs, recall that negative correlation reduces the heterogeneity in consumers' willingness to pay for bundles, because high valuations for some products in the bundles are "offset" by low valuations for other products in the bundle. However, asymmetry in variances diminishes the heterogeneity-reducing effect, because asymmetry implies that goods with low realizations and goods with high realizations (for a particular consumer) will not sufficiently offset each other. Similarly, negative correlation favors bundling only when the low valuations are still high enough to exceed marginal cost, since what is relevant to the profitability of bundling strategies is heterogeneity reduction in the *surplus* associated with a bundle, as opposed to the gross valuation.[32] If marginal costs are high relative to product valuations, then negative correlation makes it less likely that a consumer values more than one of the firm's products above cost. Thus, inducing a consumer to purchase an additional product may actually *lower* the consumer's valuation of the bundle relative to its cost.

Costs and Capacity Constraints.—Figure 5 provides a summary of the numerical experiments for three different assumptions about costs.[33] It is apparent that increasing marginal cost reduces the profitability of BSP relative to both CP and MB. This is unsurprising given that prior research has shown that increasing marginal costs tends to reduce the profit of PB relative to CP. However, there are several aspects of these results that are less obvious. First, as we explained in the above analysis of correlation, there is an interaction effect between marginal cost and correlation—increases in marginal costs are most harmful to BSP profitability when tastes are negatively correlated.

Second, we think it is remarkable how well BSP does in the cases we consider. Under the assumption of zero marginal costs, BSP is more profitable than CP in 98 percent of the experiments, and BSP is within 1 percent of the MB profits in 69 percent of the experiments. Even when marginal costs are higher—equal to half of the products' mean valuations—BSP is more profitable than CP in 83 percent of the experiments, and within 1 percent of the MB profits in 31 percent of the experiments.

Since the superiority of BSP to CP (in the profit sense) clearly depends on the level of marginal costs, one might argue that if we had focused our numerical experiments on cases with much higher costs, BSP might not have performed so well. This is surely true. In fact, in a smaller set of experiments in which we set costs equal to 80 percent of the products' mean valuations, BSP was more profitable than

---

[31] Though not shown in the figure, this reversal also occurs for PB. Crawford and Cullen (2007) show a related result: if marginal cost is high then adding a product to an existing bundle may reduce profit even when consumers' tastes are negatively correlated across products.

[32] If marginal costs are all zero and there is free disposal, the same reverse effect from correlation may also occur if there is a significant mass of negative valuations for any product. The reasoning is the same as with positive marginal costs.

[33] We exclude the experiments for positive and equal marginal costs from the figure because they add no further insight—they look very similar to the zero marginal cost case.
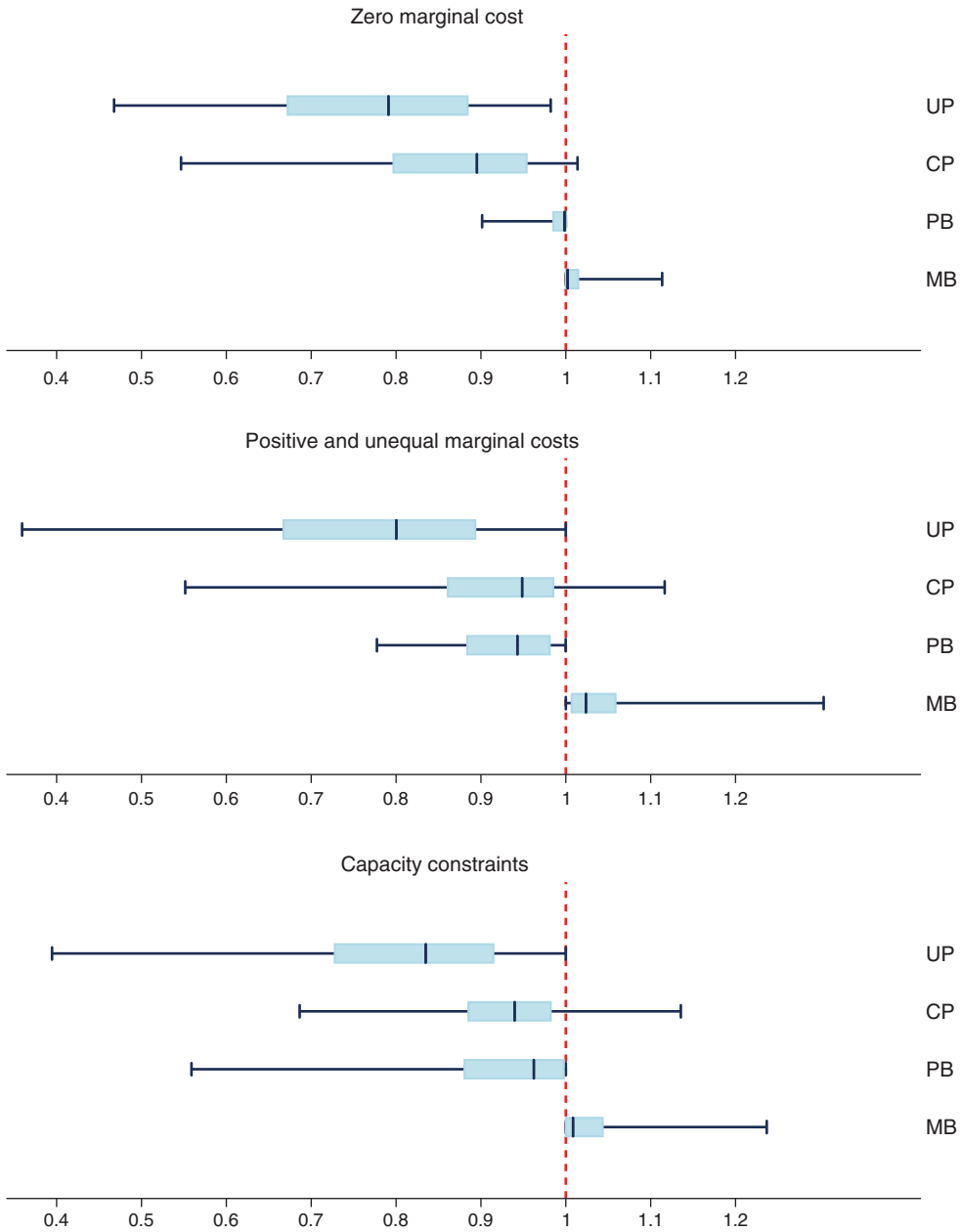
FIGURE 5. DISTRIBUTIONS OF PROFITS FOR EACH PRICING STRATEGY, RELATIVE TO BSP,
UNDER DIFFERENT ASSUMPTIONS ON MARGINAL COSTS

Note: Each box plot depicts the 1st, 25th, 50th, 75th, and 99th percentile of the distribution of profit relative to the profit from BSP.

CP in just over half of the experiments. Presumably, if we set marginal costs high enough we would find that CP is routinely more profitable than BSP. However, our chosen cost regimes are reasonable approximations of the markets where bundling strategies are a priori the most relevant (e.g., media and entertainment). For

example, based on recently reported empirical estimates, the assumption reflected in the middle panel of Figure 5 (that marginal cost is equal to half the products' mean valuations) appears to be approximately correct for cable television distributors—an industry in which bundling has recently been an important concern.[34]

A third point to note from Figure 5 is that BSP is actually less sensitive than PB to increases in marginal costs. As Adams and Yellen (1976) first argued, high marginal costs tend to reduce the profitability of bundling because individuals may purchase bundles that include a good they value less than marginal cost—violations of the so-called exclusion condition. BSP is less sensitive to this concern than PB, because BSP allows consumers to purchase bundles smaller than the full bundle. Consequently, Figure 5 clearly shows that PB is more sensitive to cost increases than BSP. Such reasoning also helps to explain why BSP tends to perform quite well even in the high cost experiments.

Figure 5 also provides a glimpse into the sensitivity of the different pricing schemes to capacity constraints.[35] In these experiments we find that BSP is more profitable than CP in 85 percent of the experiments, and BSP is within 1 percent of the MB profits in 52 percent of the experiments. An interesting feature of the experiments with capacity constraints is that BSP appears to stimulate sales of the low demand goods, above what CP tends to achieve. Specifically, in the experiments with capacity constraints, realized demand for the least popular good is typically 9 percent higher under BSP than under CP. Intuitively, low demand goods tend to be purchased as part of large-size bundles, where the discount for the bundle implies per-good prices that are less than the CP price.

Demand Asymmetry.—Intuition suggests that increasing demand asymmetry across products may favor CP over BSP, because the ability to set different prices for different goods becomes increasingly important. One simple measure of demand asymmetry is the ratio of the highest price to the lowest price under CP. The greater this ratio, the more restrictive is BSP since it requires all single-good purchases to be equally priced. Somewhat surprisingly, however, in our numerical experiments this price ratio is essentially uncorrelated with the relative profit of CP versus BSP. Even if we look at the top 10 percent of experiments in terms of demand asymmetry (as measured by the price ratio), we find that BSP is still more profitable than CP in 90 percent of these experiments. In other words, a high degree of asymmetry does not imply that CP will be more profitable than BSP.

To further understand the effects of demand asymmetry, consider an example in which there are two goods (1 and 2), each with zero marginal cost, and two consumers (A and B), with valuations as follows:

	v_1	v_2
A	2	0
B	0	1

.

---

[34] See the estimates reported in Chu (2008) and Crawford and Ali Yurukoglu (2009).

[35] We are unaware of prior papers examining the impact of capacity constraints on the relative profits of PB, CP, and MB (or BSP, which is new to the literature). A thorough analysis of this issue would include variation in the level of capacity constraints, which we leave to further research.

Good 1 has significantly greater demand than good 2, and CP earns 50 percent more profit than BSP. Demand asymmetry aside, one basic reason why BSP performs poorly in this example is that no consumer wants more than one good—bundling is irrelevant in these circumstances.

However, this example is useful for explaining the role of demand asymmetry. If the "2" in the above example were a "1," so that each product had the same demand curve, then BSP would do just as well as CP. Now suppose the "2" in the example above were a "20," giving rise to a much higher degree of demand asymmetry. In that case, CP would be only 5 percent more profitable than BSP: almost all of the profits come from selling good 1, and BSP can extract those profits by charging a high price for the bundle of size 1. As demand becomes more and more asymmetric, CP remains more profitable than BSP, but the absolute advantage (1 unit of profit) does not change, and in percentage terms BSP profits actually get closer and closer to CP profits.

This intuition carries over to settings where consumers may value (and purchase) more than one product, and also where values are continuously distributed. Suppose consumers' tastes for the firm's two products have the bivariate normal distribution with zero means and zero correlation, but with different variances.[36] When the standard deviation of the first product is three times larger than that of the second product, CP is nearly 9 percent more profitable than BSP. If the standard deviation is ten times larger, CP is only 3 percent more profitable. Thus, compared to a baseline of symmetric demand across products, some degree of demand asymmetry causes CP to be more profitable than BSP. But CP's relative advantage actually diminishes for higher levels of asymmetry.

The above example—in which each consumer has a strictly positive valuation for only one of the two goods—suggests another dimension of heterogeneity that could be relevant: consumers may vary in their taste for the *number* of underlying products. For example, a firm may be selling ten products, and some consumers may be willing to purchase only one product, others may be willing to purchase only two, and so on. It would be interesting to examine how the effectiveness of different bundling schemes depends on the average number of products consumers value positively, or on the variance across consumers in the number of products they value positively. In fact our experiments include this kind of demand heterogeneity due to the assumption of free disposal: negative values for a product are assigned zero value (so that there is no disutility from purchasing a bundle that contains a negatively valued good). Thus, if different consumers place positive value on different products (negative correlation), asymmetry across products tends to diminish the profitability of BSP.

Also note that diminishing marginal utility is similar to this form of heterogeneity, because it explicitly relates an individual's demand to the total number of products purchased. We discuss diminishing marginal utility in sections IIIA and IVD. A complete analysis of how the profitability of different bundling schemes is affected by heterogeneity in the number of products consumers value is left for future research.

---

[36] This is the same as the "normal-var" set-up in our numerical experiments and is also the basic set-up we use in our empirical model below.
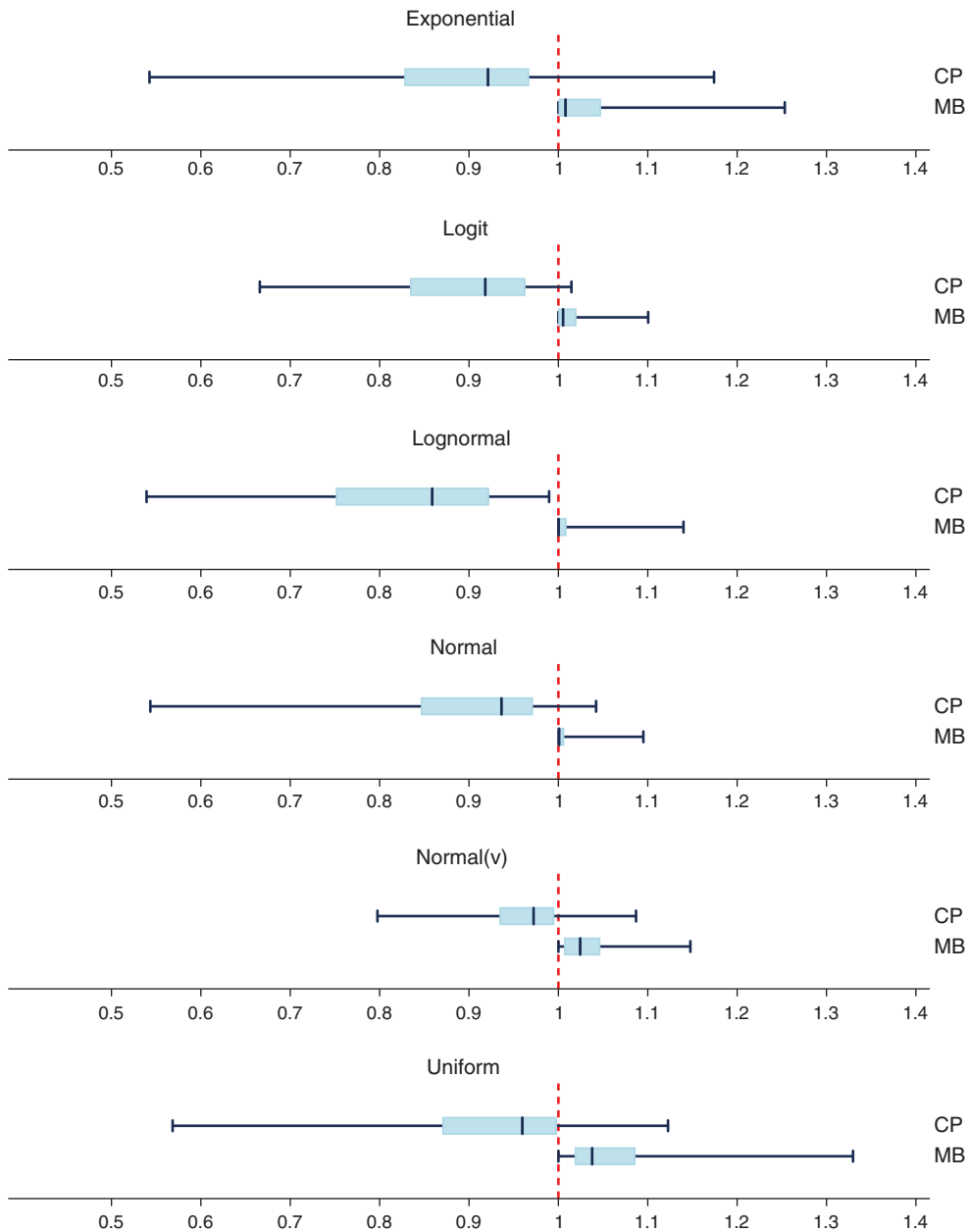
FIGURE 6. DISTRIBUTIONS OF PROFITS FOR EACH PRICING STRATEGY, RELATIVE TO BSP,
FOR DIFFERENT DISTRIBUTION FAMILIES

Note: Each box plot depicts the 1st, 25th, 50th, 75th, and 99th percentile of the distribution of profit relative to the profit from BSP.

Parametric Families.—Figure 6 shows that the choice of parametric family may not be innocuous in terms of the profitability of different pricing strategies, again highlighting the importance of considering a wide range of distributional assumptions. For instance, for the logit distribution (which is one of the most commonly used in empirical research) BSP is almost always more profitable than CP regardless

of the level of marginal costs. The same is true for lognormal distributions. Though it is not evident in the figure, we also found that the role of a given parametric family can depend on which assumption about costs is applied. For example, when marginal costs are all zero, BSP is always more profitable than CP if valuations are independent draws from the exponential distribution. However, unequal marginal costs or capacity constraints can change this.

Importantly, the normal distribution (including cases with independence, positive correlations, negative correlations, and unequal variances) seems to be the least restrictive, in the sense that either CP or BSP may be the most profitable under any assumption on costs. In the empirical example we study in Section IV, we assume normally distributed tastes.

It is natural to wonder about the underlying sources of the differences shown in Figure 6. Our analysis above highlights the characteristics of taste distributions that appear to have first-order importance for the relative profitability of BSP (e.g., correlation in tastes, demand asymmetry). However, Figure 6 suggests that the shapes of the taste distributions may also play a role. For example, standard measures like skewness and kurtosis may be useful in determining which pricing scheme is most profitable. Another relevant measure may be the fraction of potential surplus that is appropriable by a monopolist, as suggested by Michael A. Spence (1980). These are interesting issues, but we defer a thorough exploration of them to future research.

Summary of Combined Effects.—The prior literature proposed several factors that enhance the profitability of PB relative to CP:

  (i) Lower marginal costs;

 (ii) Increasing number of goods; and

(iii) More negative correlation in consumers' valuations. Symmetric demand is an obvious addition to that list (although we are unaware of any specific papers that discuss this issue). To a degree, these factors have similar effects on the relative profitability of BSP and CP, but we also find a couple of important complications:

    (a) Negative correlation in tastes enhances the profitability of BSP when the variance in consumers' tastes is roughly symmetric across goods, and consumers' tastes are well above marginal costs. Otherwise the reverse applies.

    (b) Increasing demand asymmetry initially decreases then increases the profits of BSP relative to CP.

It is natural to wonder how much of the variation in the relative profits of BSP and CP can be explained by a combination of all these factors. One simple way of addressing this question is to run regressions using our experimental outcomes. This approach is obviously imperfect, since we may not have the right functional form or the right measure of demand asymmetry. Hence, we expect any such regression

TABLE 4—REGRESSION ANALYSIS OF (BSP PROFIT / CP PROFIT)

	Exponential	Logit	Lognormal	Normal	Normal(v)	Uniform	All combined
Number of products dummies:							
$K = 3$	0.069	0.031	0.068	0.042	0.036	0.034	0.038
	(0.003)	(0.002)	(0.002)	(0.004)	(0.001)	(0.003)	(0.002)
$K = 4$	0.122	0.053	0.120	0.078	0.070	0.064	0.071
	(0.003)	(0.002)	(0.002)	(0.004)	(0.001)	(0.003)	(0.002)
$K = 5$	0.160	0.071	0.159	0.106	0.096	0.092	0.098
	(0.003)	(0.002)	(0.002)	(0.004)	(0.001)	(0.003)	(0.002)
MC scenario dummies:							
Positive, equal	−0.064	−0.009	−0.023	−0.012	−0.022	−0.039	−0.028
	(0.005)	(0.003)	(0.004)	(0.007)	(0.002)	(0.005)	(0.003)
Positive, unequal	−0.159	−0.002	−0.096	−0.010	−0.033	−0.119	−0.070
	(0.005)	(0.003)	(0.004)	(0.007)	(0.002)	(0.005)	(0.003)
Capacity constraints	−0.179	−0.027	−0.029	−0.016	−0.045	−0.110	−0.068
	(0.005)	(0.003)	(0.004)	(0.007)	(0.002)	(0.005)	(0.003)
Covariance structure dummies:							
Negative	0.504	0.237	0.550	0.370	0.063	0.499	0.325
	(0.006)	(0.003)	(0.004)	(0.008)	(0.002)	(0.006)	(0.003)
Positive	−0.148	−0.073	−0.140	−0.075	−0.053	−0.093	−0.097
	(0.005)	(0.003)	(0.004)	(0.007)	(0.002)	(0.005)	(0.003)
Cov. × MC interactions:							
Neg. × (pos/eq.)	−0.131	−0.007	0.006	−0.013	−0.046	−0.042	−0.039
	(0.006)	(0.004)	(0.005)	(0.010)	(0.002)	(0.006)	(0.004)
Neg. × (pos/uneq.)	−0.377	0.058	−0.059	0.034	−0.044	−0.142	−0.088
	(0.006)	(0.004)	(0.005)	(0.010)	(0.002)	(0.006)	(0.004)
Neg. × (cap. constr.)	−0.374	−0.099	−0.224	−0.157	−0.065	−0.286	−0.201
	(0.006)	(0.004)	(0.005)	(0.010)	(0.002)	(0.006)	(0.004)
Pos. × (pos/eq.)	0.047	0.007	0.007	0.008	0.017	0.028	0.019
	(0.006)	(0.004)	(0.005)	(0.010)	(0.002)	(0.006)	(0.004)
Pos. × (pos/uneq.)	0.104	−0.000	0.053	0.007	0.018	0.072	0.042
	(0.006)	(0.004)	(0.005)	(0.010)	(0.002)	(0.006)	(0.004)
Pos. × (cap. constr.)	0.165	0.023	0.032	0.022	0.030	0.086	0.060
	(0.006)	(0.004)	(0.005)	(0.010)	(0.002)	(0.006)	(0.004)
Asymmetry in product valuations:							
Asymmetry	−0.063	0.007	−0.082	−0.006	−0.082	−0.107	−0.033
	(0.011)	(0.005)	(0.006)	(0.008)	(0.004)	(0.005)	(0.003)
Asymmetry^2	−0.025	−0.022	0.006	−0.015	−0.017	0.032	−0.003
	(0.009)	(0.003)	(0.004)	(0.003)	(0.004)	(0.002)	(0.002)
Asymmetry × neg. corr.	−0.239	−0.068	−0.311	−0.133	0.013	−0.184	−0.094
	(0.007)	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)	(0.002)
Constant	1.171	1.077	1.189	1.074	1.053	1.139	1.115
	(0.004)	(0.003)	(0.003)	(0.006)	(0.001)	(0.004)	(0.002)
N	10,704	10,704	10,704	10,704	10,704	10,704	64,224
R^2	0.790	0.790	0.875	0.492	0.701	0.744	0.477

Notes: The dependent variable in each regression is the ratio of BSP profit to CP profit. An observation is a numerical experiment (i.e., a taste distribution with a specific set of parameters). Standard errors are in parentheses. Asymmetry is measured as the difference between the maximum and the mean of the products' mean valuations.

to explain far less than 100 percent of the variation in relative profits. Nevertheless, we report some regression results in Table 4 as a convenient way of summarizing the patterns we observe in the numerical experiments. The dependent variable is the ratio of BSP profit to CP profit, and the explanatory variables are characteristics of the taste distribution and costs. We estimate the regressions separately for each parametric family (as we expect the coefficients to vary across parametric families), as well as for the pooled set of all experiments. Demand asymmetry is measured as the

difference between the maximum and the mean of the products' mean valuations.[37] An observation is a single numerical experiment.

As expected, we find that the profitability of BSP relative to CP increases with K and decreases with marginal costs. Asymmetry favors CP, and the relationship is sometimes nonmonotonic: the coefficients on the asymmetry measure and its square have opposite signs in some cases. Also consistent with our explanation above, the interaction of the negative correlation dummy with the cost dummies indicates that negative correlation favors BSP, but significantly less so when marginal costs are high. Asymmetry combined with negative correlation is also unfavorable for BSP on average. Notably, the R^2 from these regressions is high in some cases, indicating that for some parametric families we can explain a large fraction of the variability in BSP's relative profitability using a simple, intuitive set of covariates. Moreover, these covariates are important. For example, in the uniform case, if the only covariates are the number of products and cost dummies, we obtain an $R^2 = 0.15$. Adding variables for correlation and demand asymmetry increases explanatory power to $R^2 = 0.57$. And adding the interaction terms (i.e., the full set of covariates shown in Table 4) yields an $R^2 = 0.74$.

Welfare Analysis.—As a final point on the numerical experiments, Table 5 describes the differences in social surplus resulting from the different pricing strategies. BSP and MB tend to yield significantly higher total output and higher profits (as we have seen in the previous tables). The table also shows that BSP and MB tend to reduce the deadweight loss by significant amounts, relative to CP. Interestingly, the table indicates that BSP and MB tend to also result in lower consumer surplus than CP. In our experiments, apparently BSP and MB are more like perfect price discrimination. This comes from the heterogeneity-reduction effect: there is less heterogeneity in consumers' valuations for bundles of multiple goods than there is for individual goods.

III. Estimation of Joint Distribution of Consumers' Valuations

An obvious limitation of the numerical experiments in Section III is that we cannot be certain our results will transcend the particular parameter values we covered. For this reason, the second component of our analysis utilizes an estimated model, based on data from a theater company that offers an eight-play season. We use our estimates to compute the profitability of each pricing strategy, allowing us to demonstrate that our findings apply to an empirically relevant model.

Several features make our particular empirical example an appealing context in which to study multiproduct pricing. First, the plays differ in their overall popularity, making it plausible that CP would be a sensible pricing strategy. Second, many consumers attend more than one play, making it plausible that bundling strategies may also be profitable. Third, individuals do not consume multiple units of the same play. Fourth, the assumption of no demand or cost interdependencies is reasonable. Fifth, we

---

[37] That is, if μ_k is the mean valuation for product k, we calculate $\max_k\{\mu_k\} - 1/K \sum \mu_k$. Among several measures we tried, this gave the best fit. Other measures (e.g., the range or standard deviation of μ_ks) gave qualitatively similar results.

TABLE 5—AVERAGE WELFARE EFFECTS

	CP	PB	BSP	MB
$K = 3$				
Total output	1.213	1.650	1.456	1.501
Consumer surplus	0.726	0.616	0.617	0.602
Producer surplus	1.276	1.416	1.471	1.499
Total surplus	2.002	2.032	2.087	2.101
Deadweight loss	0.499	0.468	0.413	0.400
$K = 4$				
Total output	1.625	2.285	1.987	2.044
Consumer surplus	0.980	0.805	0.807	0.770
Producer surplus	1.726	1.948	2.027	2.065
Total surplus	2.707	2.753	2.834	2.835
Deadweight loss	0.661	0.615	0.534	0.533
$K = 5$				
Total output	2.033	2.927	2.510	2.590
Consumer surplus	1.231	0.984	0.987	0.920
Producer surplus	2.174	2.486	2.591	2.641
Total surplus	3.405	3.470	3.578	3.561
Deadweight loss	0.826	0.761	0.653	0.670

Notes: Total output is calculated as the number of units sold of all K products combined. The cells report averages taken across experiments.

are confident there is no significant resale activity—these are plays produced by a small theater company, not rock concerts or professional sporting events.[38] For all of these reasons, our empirical example is a remarkably clean setting, in which we can abstract from the same complicating factors that theoretical analyses of bundling typically do.

There is an emerging empirical literature about bundling. Two studies in the marketing literature use survey response data to estimate demand and compare profits from UP, PB, and MB: R. Venkatesh and Vijay Mahajan (1993) and Kamel Jedidi, Sharan Jagpal, and Puneet Manchanda (2003). Crawford (2008) is the first to study bundling with market-generated data. Crawford tests the hypothesis that demand for a bundle of cable channels becomes less heterogeneous as more channels are added to the bundle, which he finds to be the case.[39] A number of recent papers also estimate the welfare effects of bundling in various settings, as we do here: Dmitri Byzalov (2008) and Gregory S. Crawford and Yurukoglu (2009) study bundling in cable television; Justin Ho, Katherine Ho, and Julie Holland Mortimer (2008) study full-line forcing in video rental markets; and Ben Shiller and Joel Waldfogel (2009) examine bundling in music sales. Several of the studies assess BSP (citing this study), all finding it to be more profitable than CP.

A. *Data Summary*

The data for our empirical analysis come from *TheatreWorks*, a theater company based in Palo Alto, California. We observe all ticket sales for *TheatreWorks*' 2003–2004 season, which consisted of 229 performances of eight different plays or

---

[38] See Leslie and Sorensen (2009) for an empirical analysis of ticket resale.
[39] Based on a calibrated demand model, Crawford (2008) also argues that adding a top-15 cable channel to a bundle and re-optimizing prices leads to 5.5 percent lower consumer surplus and 6.0 percent higher profit.

Play	Type	Number of performances	Average attendance	Ticket sales (subscription)	Ticket sales (nonsubscription)
A Little Night Music	Musical	30	294.87	7,018	1,828
All My Sons	Drama	33	233.85	6,826	891
Bat Boy	Musical	30	263.93	6,782	1,136
Memphis	Musical	30	352.40	6,999	3,573
My Antonia	Drama	26	312.38	7,002	1,120
Nickel and Dimed	Drama	26	343.62	6,800	2,134
Proof	Drama	25	319.88	6,885	1,112
The Fourth Wall	Comedy	29	313.83	7,385	1,716
Total		229	302.21	55,697	13,510

Notes: Three plays (*Bat Boy*, *All My Sons*, and *The Fourth Wall*) were performed at the Lucie Stern Theater in Palo Alto (capacity = 428). The remaining five were performed at the Mountain View Center for the Performing Arts (capacity = 589).

musicals. Table 6 provides summary information for each of the eight plays. A total of 69,207 tickets were sold for the eight plays.

Consumers could purchase tickets to individual plays at a uniform price, but most of the tickets (80 percent) were purchased as part of a subscription. *TheatreWorks* offered three subscription packages:

  (i) The full eight-play season;

  (ii) Any combination of five plays; or

  (iii) A prespecified bundle of three plays.[40]

These subscriptions were offered at discounted prices, in the sense that the per-play price was significantly lower for subscriptions than for ordinary box office sales for individual plays.

Table 7 summarizes the purchase options and their average prices.[41] There were 5,139 subscribers to the eight-play bundle, 2,794 subscribers to a five-play bundle, and 205 subscribers to the three-play bundle. The popularity of the flexible five-play subscription is a particularly important feature of the data. Observing which five plays these subscribers selected allows us to identify the covariance of tastes across plays—e.g., if we observe that two plays tend to be included together disproportionately often in the five-play combination, we know that tastes for those two plays are more positively correlated. Conversely, if another pair of plays is rarely included in the same bundle, we can infer that tastes for those two plays are less positively correlated. This type of information is crucial to our analysis. If we had data only on

[40] The prespecified bundle consisted of the only three plays that were performed at *TheatreWorks'* secondary venue, a smaller theater in Palo Alto, CA.

[41] In fact prices also vary by time of week (but not by play). We therefore report simple (unweighted) averages of these prices. Note also, prices do not vary by seat quality. This is because the venues are small enough that the variation in seat quality is fairly minor.

Table 7—Sales by Purchase Option

Purchase option	Price per play ($)	Number of consumers
Nonsubscription:		
1 play	40.80	8,131
2 plays	40.80	1,409
3 plays	40.80	555
4 plays	40.80	224
Subscription:		
3-play bundle	36.20	205
5-play pick	37.00	2,794
8-play bundle	34.55	5,139

Notes: For nonsubscription purchases, the numbers of consumers in each purchase option are computed by extrapolating the purchase patterns of the consumers whose identities we could observe to the full sample of nonsubscription purchases. See text for an explanation. The 3-play subscription bundle was for the specific 3 plays performed at the (smaller) Lucie Stern Theater in Palo Alto, which is why the per-play price is lower than the 5-play bundle. Consumers purchasing the 5-play subscription could combine any 5 plays of their choice.

aggregate sales for each play we would be unable to identify the covariance structure of demand.[42]

In fact, the empirical correlations in our data make intuitive sense. For example, the five-play subscriptions disproportionately include *Bat Boy*, described in the brochure as a "wacky new musical," together with *Memphis*, described as a "rafter-rattling musical comedy." Conversely, tastes for *Bat Boy* appear to be negatively correlated with *All My Sons*, a classic Arthur Miller drama billed as an "intense, compelling tale of love, greed, and personal responsibility," because this pairing of plays is relatively uncommon in the five-play subscriptions.

B. *Empirical Model*

We now explain how we estimate the joint distribution of consumers' valuations. It is important that such an approach allow for nonzero covariances in tastes, because covariance is a major determinant of the relative profits from different bundling-type schemes, and that we allow for consumers to purchase multiple products.[43] Our empirical specification is based on an underlying model of individual consumer utility maximization and follows the approach in the theoretical literature on bundling. The firm offers $j = 1, \ldots, J - 1$ bundles containing combinations of the $k = 1, \ldots, K$ products. There is also a Jth option for consumers, which is the outside alternative. We assume the net utility to consumer i from option j is given by

$$u_{ij} = \begin{cases} V_i'D_j - \alpha p_j & : j = \{1, \ldots, J - 1\} \\ 0 & : j = J \end{cases},$$

[42] Steve Berry, James Levinsohn, and Ariel Pakes (2004) utilize a similar identification strategy in their study of demand for cars, in which they exploit second-choice data to help identify cross-price elasticities.

[43] Crawford and Yurukoglu (2009) also estimate the correlation structure in preferences for bundle components. Prior papers that also incorporate multiple purchases by individuals include Jean-Pierre Dubé (2004) and Igal Hendel (1999).

where V_i is a $K \times 1$ vector of valuations for the individual plays, D_j is a $K \times 1$ vector of indicators for whether each play is included in bundle j, p_j is the price of the bundle, and $\alpha > 0$ measures the sensitivity to price. As always in the bundling literature, we assume there are no demand-side complementarities from purchasing particular plays together.

We allow for two classes of consumers: theater lovers and regular consumers. In fact the data support this description.[44] Formally, we assume that consumers' product valuations are distributed according to a K-dimensional bimodal normal distribution, with censoring at zero to incorporate free disposal:

$$V_i = \max\{\theta_i + \epsilon_i, 0\}, \quad \text{where}$$

$$\theta_i = \begin{cases} \bar{\theta} & : \text{ probability } \lambda \\ 0 & : \text{ probability } (1 - \lambda); \quad \text{and} \end{cases}$$

$$\epsilon_i \sim N(\mu, \Sigma).$$

In this notation, μ is a $K \times 1$ vector of means, Σ is a $K \times K$ variance-covariance matrix, and $\bar{\theta}$ is a scalar additive component (equal for all plays). A fraction λ of consumers are theater lovers, for whom the marginal distribution of play valuations is shifted upward by some amount $\bar{\theta}$ that is constant across plays. A fraction $(1 - \lambda)$ are regular consumers with no particular preference for seeing plays in general.

The conditional means of V are not well identified separately from the variances. Intuitively, increasing the variance in valuations for a particular play and increasing the mean of the valuations for that play both lead to higher demand for the play. To address this, in our base specification we impose the restriction that all mean terms equal zero ($\mu(k) = 0, \forall k$)[45] but leave the variance-covariance matrix unconstrained.[46] We also estimated the model based on the restriction that all variances equal 1, with the mean terms unconstrained, but we found that version to be too restrictive in the following sense: BSP is always more profitable than CP, even in counterfactuals where we dramatically increase the asymmetry across products by making the mean valuations for each play very different across plays. In contrast, in the specification with free variances it is possible that either CP or BSP may be more profitable, depending on the particular values of the variance terms. Given the objective of the empirical analysis, we viewed this as a desirable attribute for the model. Note that our approach has the implication that a high-quality play will have a higher variance in consumers' valuations—i.e., our model captures quality via the variance terms rather than the means, which is unconventional in the literature.

The season of eight plays implies 255 possible product combinations. This includes each individual play, the preset bundle of three, the full bundle of all eight,

---

[44] As we explain in an the Web Appendix.

[45] We also estimated a slightly less restrictive specification that allows the mean terms to differ from zero but constrains the means to be equal across plays ($\mu(k) = \mu, \forall k$). However, the point estimate of μ in that model is 0.0255 and is not statistically different from zero. In what follows we simply focus on the model that restricts μ to be zero.

[46] We also normalize the variance of valuations for play (1) to equal 1: $\Sigma(1,1) = 1$. This is necessary for the scaling of ϵ to be separately identified from the price sensitivity (α).

56 possible combinations of five plays (for *pick*-5 subscribers), and any other combination by consumers adding individual plays. In fact we observe zero sales of bundles of six or seven plays. We therefore exclude these combinations from the consumer's choice set to reduce the computational burden. Hence, we model the demand for 219 different bundles, plus an outside alternative, giving a total of 220 possible choices (i.e., $J = 220$). Note that capacity constraints are infrequently binding in the data—only 27 of the 229 performances were sold out—leading us to abstract from their impact in the estimation. In the subsequent counterfactual analyses we check whether capacity constraints are binding.

For nonsubscription purchases we cannot always determine whether the individual purchased multiple plays, because roughly half of these purchases were made anonymously at the box office. This means that we do not observe market shares for combinations involving fewer than five plays purchased by the same individual. For this reason, we estimate the model's parameters by the method of simulated moments (see McFadden 1989, and Pakes and David Pollard 1989). Using a method of moments approach allows us to treat this data problem conservatively, without throwing away information that we *do* have from nonsubscription purchases. Specifically, we use only moment conditions that are based on market shares we directly observe:

- Share of consumers who chose all eight plays (one moment condition);
- Shares of consumers choosing specific combinations of five plays (56 moment conditions);
- Share of consumers choosing the preset bundle of three plays (one moment condition);
- Overall market shares of each play: i.e., what fraction of consumers purchased a given play as part of *any* bundle (eight moment conditions).

The last set of moment conditions utilizes information from nonsubscribers without imposing any assumptions about their pattern of multiplay purchases.[47]

To ensure that the estimated demand model yields predicted prices that are close to the observed prices, we impose a supply-side pricing constraint in the estimation.[48] For any given set of parameters of the above demand system, we can compute the profit-maximizing prices under the actual *TheatreWorks* pricing structure: a price for any individual play, a price for the preset bundle of three, a price for choosing any five, and a price for all eight plays.[49] Solving for these prices for each iteration of conjectured parameters is computationally burdensome, however, so we simplify the constraint in the following way. Rather than jointly optimize all four prices in the *TheatreWorks* pricing scheme, we jointly optimize the price of any individual play and the price of all eight plays. We then "fill in" the three-play and five-play prices

by assuming their ratios to the single-play price are equal to the ratios actually set by *TheatreWorks*. This reduces the number of prices we must optimize from four to two, which we found to be essential for computational feasibility. Hence, the specific constraint we impose is that the predicted single-play price, and the predicted price for the subscription to all eight plays, are equal to the observed prices.[50]

To compute the market shares, we must know the market size, M. Usually researchers choose the market size based on some information about the number of potential consumers. In our case, an additional benefit to utilizing a supply-side pricing constraint is that we can estimate the market size instead of assuming some value for it.[51] Including α, Σ, $\overline{\theta}$, λ, and M, we estimate a total of 39 parameters. Let Θ denote the set of parameters to be estimated. For a given set of parameters, Θ, we draw n_s simulated consumers based on the above distribution of product valuations, compute the optimal bundle choice for each simulated consumer, and compute optimal prices. The estimator chooses the parameters Θ to match the market shares among the simulated consumers to the market shares we observe in the data, conditional on predicted prices being equal to actual prices.

More formally, let $\tilde{s}_l$ and s_l denote the simulated and observed market shares, respectively, for purchase option l. Let $\tilde{p}_1(\Theta)$ denote the implied optimal single-play price for given parameters Θ, and let p_1 denote the actual single-play price. Similarly, let $\tilde{p}_8(\Theta)$ and p_8 denote the implied and actual full season subscription prices. Define $\tilde{p}$ and p as the stacked vectors of predicted and observed prices. We construct moment conditions of the form $m_l(\Theta) = \tilde{s}_l(\Theta) - s_l$ and select Θ to minimize $m'Wm$ subject to the constraint $\tilde{p} = p$, where m is the stacked vector of moment conditions, and W is a weighting matrix.

Finally, it is important to discuss what variation in the data identifies each parameter in the model. Due to space constraints we provide a detailed discussion of identification in the Web Appendix (available online). Briefly, however, note that variances are identified by relative shares of each play; covariances are identified by bundle choices of *pick*-5 subscribers; and price sensitivity is identified from variation in per-play prices.

C. *Results*

Parameter estimates for the structural demand model are presented in Table 8A.[52] The variance coefficients from the distribution of ϵ vary from 1.00 to 3.27. The estimates for the covariances of ϵ vary from 0.79 to 2.55. It is important to note that Σ is the covariance matrix of ϵ. That is to say, Σ captures the correlation structure conditional on being a theater lover, or conditional on not being a theater lover. However, the correlation structure of the unconditional distribution of play valuations, V, also depends on the probability of being a theater lover and the increment in utility for these consumers. Intuitively, taste correlations should be even more positive for the

---

[50] In practice, these conditions hold with equality within machine precision.
[51] As mentioned below, in the Web Appendix we explain the identification of our model, including how the price constraint identifies the market size.
[52] Standard errors for the variance-covariance matrix are reported in Table 8B.

TABLE 8A—ESTIMATED COEFFICIENTS

| | Covariances (Σ) | | | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	1.0000							
(2)	0.9357	1.2200						
(3)	1.2125	1.4150	1.7208					
(4)	0.9793	1.3381	1.4859	3.2685				
(5)	0.8743	1.1055	1.2207	1.7920	1.4308			
(6)	1.1602	1.3451	1.6211	1.9090	1.3801	1.9610		
(7)	0.7886	1.0600	1.2199	1.7924	1.2127	1.4517	1.5086	
(8)	1.1133	1.2873	1.5878	2.5509	1.5597	1.9529	2.0171	3.0732

	Estimate	SE
Price sensitivity (α)	4.5937	(0.0851)
Probability of theater lover (λ)	0.0805	(0.0060)
Increment for theater lovers ($\bar{\theta}$)	2.0561	(0.1665)
Market size	36,055	(972)

Notes: Standard errors for Σ are in Table 8B. All parameter estimates are significant at the 1 percent level.

TABLE 8B—STANDARD ERRORS FOR ESTIMATED COVARIANCES (Σ) IN TABLE 8A

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(2)	0.0568	0.0237						
(3)	0.0257	0.0575	0.0232					
(4)	0.0752	0.0675	0.1515	0.0774				
(5)	0.0325	0.0526	0.0459	0.0735	0.0310			
(6)	0.0293	0.0506	0.0449	0.0629	0.0719	0.0338		
(7)	0.0344	0.0448	0.0432	0.0633	0.0490	0.0824	0.0274	
(8)	0.0538	0.0507	0.0624	0.1004	0.0530	0.0618	0.0948	0.0581

unconditional distribution, because theater lovers have a positive shift in the valuations of *all* plays.

When we compute pairwise correlation coefficients for the unconditional distribution of play valuations, we find that all correlations lie between 0.60 and 0.98 (the mean correlation coefficient is 0.81). This is important because positive correlation in the demand system tends to reduce the profitability of bundling-type strategies relative to component pricing. We return to this issue in the next subsection on counterfactual pricing experiments.

The estimated probability of an individual's being a theater lover is 0.081, and the estimated market size is 36,055. We estimate that theater lovers' utility for any single play is higher than for regular consumers by an amount equal to 2.06 times the standard deviation of the conditional valuation of play 1 (*A Little Night Music*), which is normalized to 1. The large magnitude of the increment to utility for theater lovers suggests that large-sized bundles are disproportionately purchased by theater lovers. Indeed, this is true. Our estimated demand model predicts that 63 percent of non–theater lovers choose the outside option, while the predicted proportions choosing 1, 2, 3, or 4 individual goods are 10 percent, 5.5 percent, 2.2 percent, and 0.28 percent. Among this same group of consumers, the predicted market shares for the preset three-play bundle, the *pick*-5 bundle, and the all-eight bundle are 0.34 percent, 8.9 percent, and 9.9 percent. For theater lovers, on the other hand,

only 9.0 percent choose the outside good, while 14 percent choose a bundle of five (accounting for 12 percent of *pick*-5 purchases), and 66 percent subscribe to all eight plays (accounting for 37 percent of all-eight purchases).

The nonmonotonicity of predicted market shares with respect to bundle size, even among the set of non–theater lovers, is consistent with the high degree of correlation in the estimated distribution of tastes for individual plays (both unconditionally as well as conditionally on theater-lover status): consumers tend to like either most of the plays, or none at all. However, the model requires the presence of theater lovers to explain why the observed market shares decline relatively gradually with respect to size in the lower range of bundle size while rising abruptly for the largest bundles.

Overall, the model fits the data fairly well. Even though the individual plays' market shares vary from 20 percent to 30 percent, the predicted market shares from our model are all within 1.1 percentage points of the actual shares. We also find a reasonably close fit between the actual and predicted market shares for different bundle sizes, although we slightly overpredict the fraction of subscription purchases. The actual fraction that choose to be *pick*-5 subscribers is 5.5 percent, and we predict 9.3 percent. The actual fraction of all-eight subscribers is 10.0 percent, and we predict 14.5 percent.

D. *Analysis of Alternative Pricing Strategies*

In this subsection we compare the profitability of the various pricing schemes in the context of our estimated demand model. We also compute the impact of consumer surplus from each pricing strategy. This is interesting because bundling, like price discrimination more generally, has ambiguous effects on consumer welfare relative to uniform pricing.[53] There are some interesting perturbations to the model that we also explore, such as diminishing marginal utility.

Counterfactual Pricing Analysis.—Using the estimated demand model, we compute profits and consumer surplus under each of UP, PB, CP, BSP, and MB. We also compute the profit associated with the pricing scheme actually implemented by *TheatreWorks*, referred to as TW. Under TW, the firm sets a uniform price for each play, a discount for one particular three-play bundle, a discount for choosing any five plays (*pick*-5), and a discount for the bundle of all eight plays. In our baseline model we assume zero marginal costs and no capacity constraints, which seems reasonable given how few performances sold out. Below, we examine how capacity constraints would affect the relative profits of the different pricing strategies.

Recall that in the estimation we impose a supply-side pricing constraint based on two of the four prices under the TW scheme: i.e., the single-play price and the price for all eight plays. In estimation, these two predicted optimal prices exactly match the observed prices (by construction). However, in the TW counterfactual we jointly optimize all four prices. Hence, we expect the TW counterfactual prices to be close to the actual prices, but not necessarily equal.

---

[53] See Leslie (2004) for a similar empirical analysis of the welfare effects of price discrimination, which also happens to be in the context of theater ticket pricing.

TABLE 9—COUNTERFACTUAL PRICING

	UP	PB	TW	CP	BSP	MB
p_1	35.60		44.55	27.79	56.41	48.25
p_2				30.07	46.92	43.08
p_3			38.01	34.67	41.12	40.57
p_4				44.08	37.72	38.68
p_5			36.68	31.46	36.80	38.11
p_6				38.89	35.04	36.54
p_7				33.23	34.01	35.23
p_8		30.81	33.30	37.90	32.89	34.29
Revenue	66.85	63.67	67.57	67.81	68.42	69.50
CS	55.03	54.37	54.02	55.88	54.75	52.62

Notes: For UP, p_1 is the optimal uniform price for a single play. For PB, p_8 is the optimal per-play price for the bundle of all eight plays. TW is the pricing scheme currently employed by the theater company: p_1 is the single-play price, p_3 is the per-play price for a specific bundle of three plays, p_5 is the per-play price for any combination of five plays, and p_8 is the per-play price if you buy all eight. For CP, p_1–p_8 are the prices for the eight individual plays, and for BSP, p_1–p_8 are the per-play prices for any bundle containing the corresponding number of plays. For MB, p_1–p_8 are mean per-play prices for bundles of a given size (e.g., p_1 is the mean single-play price, p_2 is the mean price for all two-play bundles, and so forth). The revenue and consumer surplus numbers are normalized by the market size—i.e., we report revenue per consumer.

Table 9 summarizes the results. The interpretation of the prices $(p_1,...,p_8)$ varies across regimes, as explained in the note to the table. The revenue and consumer surplus (CS) results are normalized by the market size (i.e., figures are per consumer). Profits from the different pricing schemes vary from 63.67 under PB to 69.50 under MB (a difference of 9.2 percent). In this case the variability in profits across price structures is somewhat low compared to many of the simulations in Section III. Nevertheless, it is clear that the choice of price structure can be an important decision.

It is interesting that PB is the least profitable of the pricing strategies we examine. Bakos and Brynjolfsson (1999) and Fang and Norman (2006) show that PB becomes more profitable (relative to CP) as the number of goods increases (with zero marginal cost). But in this example with eight goods, PB performs quite badly. Even UP is more profitable than PB in this setting (by 5 percent). This reinforces the point that PB is not necessarily a good option for firms.

Focusing on BSP in Table 9, we find that:

 (i) BSP attains 0.9 percent higher profit than CP; and

 (ii) BSP attains 98.5 percent of the profit from MB.

These results are striking for a couple of reasons. MB requires the firm to set 255 distinct prices in this example, while BSP involves only eight prices. It is also important to note that our empirical example happened to yield an estimated demand system that is somewhat unfavorable to bundling-like strategies. We find a very high degree of positive correlation in valuations—all correlations lie between 0.60 and 0.97. The fact that BSP is more profitable than CP in this setting is interesting, even if the differences are not economically large.

Under BSP the price per play varies from $56.41 (for one play) to $32.89 (for all eight plays)—a discount of 42 percent on the single play price for full season subscribers. Note also that under BSP the price for seeing a single play ($56.41) is greater than the maximum price for any play under CP ($44.08). As explained in the previous section, BSP encourages consumers to purchase multiple plays by a combination of raising the price for one play and lowering prices for multiple plays. Under CP 9.9 percent of consumers attend exactly one play, while under BSP only 2.6 percent of consumers attend just one play. Under CP 9.0 percent of consumers attend all eight plays, while under BSP 12.6 percent do so.

It is interesting to compare the performance of CP and BSP in relation to the highest-demand play. One tends to expect that CP will generate more profit from these kinds of products than would BSP, although we have argued and demonstrated that BSP is also effective at extracting surplus in the presence of asymmetric demand. The play with the highest demand is play 4—under UP play 4 has the highest level of sales, and under CP play 4 has the highest optimal price. For CP we compute that 25.2 percent of consumers attend play 4, and every ticket is sold at the profit-maximizing price of $44.08. Under BSP we find that 27.1 percent of consumers attend play 4. The average (per play) price paid by consumers that attend play 4 under BSP is $36.70. It follows that BSP obtains 21 percent less revenue from play 4 than does CP.[54] Since BSP attains higher overall revenue, it must be that BSP extracts more surplus than CP for the lower demand plays. For example, we find that BSP yields 7 percent more revenue for play 1 (the lowest-demand good) than CP.

As expected, we predict optimal prices under TW that are very close to the actual prices set by *TheatreWorks*.[55] The predicted single-play price under the TW scheme is slightly higher than the observed price set by *TheatreWorks*, and the predicted full-season subscription price is slightly lower than the actual price. Thus, our estimated model indicates more aggressive discounting than *TheatreWorks*' actual price schedule, but not by much. The TW price structure appears to perform quite well in the counterfactuals. As shown in the table, the profit under TW is marginally less than for CP, despite the fact that TW involves half the number of prices as CP. This reinforces the value of bundling-like strategies, since TW incorporates a degree of bundling into its structure.

Model Perturbations.—Given our estimates of demand, it is clear that BSP is the superior pricing strategy among the simple alternatives we consider. To evaluate the robustness of this conclusion, we ask how we would have to change the demand system to reverse the conclusion.

We have emphasized throughout this study that BSP is able to perform well even in the presence of highly asymmetric demand. As a measure of asymmetry, in the baseline model above, the highest price for a play under CP ($44) is almost 60 percent greater than the lowest price ($28). But what if we amplified this difference? How much would we have to exacerbate the differences in plays' qualities to make CP more profitable than BSP? To examine this question, we took min-preserving

---

[54] Revenue is synonymous with variable profit in this context, because the marginal cost of each ticket is zero. Since we have no information on fixed costs we do not refer to these numbers as profits.

[55] To understand why optimal prices are not exactly equal to actual prices, recall that in the estimation, we impose the optimality of only the individual-play and all-eight prices, rather than the full set of prices.

spreads of the estimated variances (holding all other parameters fixed) and recomputed the optimal prices and profits under the various pricing strategies.[56] We find that BSP remains more profitable than CP even when the highest price for a play under CP ($95) is 340 percent greater than the lowest-price play ($28). However, if the price difference increases to over 385 percent (price range of $28 to $108) then CP attains higher profit than BSP. Hence, increasing demand asymmetry favors CP, but it takes a remarkably high degree of asymmetry for CP to be more profitable than BSP.

As explained above in the context of the numerical experiments, positive marginal costs should typically favor CP over BSP. If we recompute optimal prices based on the estimated demand model, assuming positive marginal costs, we can indeed get CP to be more profitable than BSP. However, the required level of marginal cost is extremely high: only when we set marginal cost as high as $40 does CP become more profitable than BSP. We suspect this is due to the high degree of positive correlation in consumers' tastes, for the following reason. Positive marginal costs tend to be bad for bundling strategies because consumers who purchase bundles may end up consuming products they value below cost, shrinking the extractable surplus from bundles. But when valuations are highly positively correlated, such violations of the "exclusion" condition will be relatively rare. It would be interesting to explore the combined effects of marginal costs and correlations in tastes in more detail, but for the present purposes we simply note that BSP's superiority over CP is remarkably robust to increases in marginal cost.

While it is true that we can make CP more profitable than BSP by significantly amplifying demand asymmetries or by imposing relatively tight capacity constraints, note that our estimated demand model exhibits a high degree of positive correlation in tastes. To highlight the degree to which positive correlation is disadvantageous to BSP, we recompute optimal prices with all the estimated covariances set to zero, holding fixed all other estimated parameters. In this case, BSP is a dramatic 20.5 percent more profitable than CP.

Finally, we argued in Section III that the inclusion of diminishing marginal utility may reduce the profit from CP by even more than it does for BSP. To verify this claim we generalize the utility function in the demand model in the following way:

$$u_{ij} = \begin{cases} V_i' D_j n_j^{\gamma} - \alpha p_j & : \quad j = \{1,\dots,J\} \\ 0 & : \quad j = J \end{cases},$$

where n_j equals the number of goods in bundle j and γ is a parameter. We set $\gamma = -0.2$ to capture diminishing marginal utility and compute optimal prices holding all other parameters fixed at the estimated values under the baseline model.[57] Unsurprisingly the profits under all pricing schemes are lowered relative to the baseline. But now BSP attains 8.8 percent higher profit than CP, compared to 0.9 percent in the baseline model. Hence, the inclusion of diminishing marginal utility can *increase* the profits of BSP relative to CP.

---

[56] That is, we hold the variance of the lowest-variance play at the estimated value, $\min[\hat{\Sigma}(k,k)]$, and increase the remaining variance terms such that they differ from $\min[\hat{\Sigma}(k,k)]$ by Δ times the corresponding differences in the actual estimates. At the same time, we inflate the covariances such that the correlations remain the same as in the actual estimates.

[57] Implicitly, $\gamma = 0$ in the baseline model.

IV. Conclusion

We have examined the profitability of several incomplex pricing strategies for multiproduct monopolists, relative to the impractical ideal of mixed bundling. Rather than focus on a simplified and unrealistic model of demand, we have relied on computational methods to explore these issues in a wide variety of demand and cost scenarios. The analysis yields two main findings. First, BSP tends to attain nearly the same level of profits as MB in a broad range of demand and cost scenarios. Hence, MB involves considerable redundancy—it includes many prices that are of negligible importance to profitability. Second, BSP tends to be more profitable than CP, even in circumstances with a high degree of demand asymmetry across products, and even when tastes are positively correlated across products.

To illustrate the empirical relevance of our findings, we estimate the demand facing a theater company that produces a season of eight plays and compute the profitability of each pricing scheme. We find that bundle-size pricing is 0.9 percent more profitable than component pricing, and bundle-size pricing attains 98.5 percent of the mixed bundling profits. Since the estimated demand model exhibits a very high degree of positive correlation, these results may understate the gains from BSP in other settings. Arguably, a limitation of our empirical analysis is that it concerns a fairly narrow setting. However, we see the simplicity of our example as a virtue: "bigger" examples invariably involve additional complexities (such as active resale markets, a much larger number of products, etc.) that make a clean empirical analysis infeasible.

The central theme of our findings is that bundling-based pricing schemes are often more profitable than component pricing. This is interesting because economists are prone to criticize firms for the lack of component pricing (e.g., movie cinemas). In fact, the appeal of bundling over component pricing is reflected in the pricing of some notable multiproduct firms. Major league baseball teams, for example, tend to employ bundling strategies (such as discounts for purchasing any nine games) more often than they employ component pricing strategies (such as charging prices that vary by opponent or by day of the week).[58] Also, online music sellers almost never charge different prices for different music tracks, even though demand is dramatically stronger for some songs than for others. But music is sold via subscription (a strategy akin to pure bundling) by at least two of the major online music stores. And while television service providers typically do not charge different prices for different channels, some offer discounts that depend on the number of channels selected.[59]

Our results represent a significant push toward understanding the merits of feasible pricing schemes for multiproduct firms. What insight does the prior literature on bundling have for a firm with five products, say? A narrow reading of the literature would imply the firm should implement mixed bundling with 31 prices, which is unlikely to be practical for most firms. A broader interpretation of the literature would suggest the firm should consider some form of bundling—which is a powerful insight—but it is unclear exactly what form that should be. This paper

---

[58] We examined the pricing for all 30 major league teams during the 2006 season. Sixteen teams employed some form of bundling (not including season-ticket subscriptions), whereas only seven charged prices that varied by opponent or by day of the week.

[59] See British Sky Broadcasting for a clear example: www.sky.com/portal/site/skycom/products/packages.

suggests specific advice to such a firm; bundle-size pricing (five prices) tends to attain nearly the same profits as mixed bundling and is almost certainly more profitable than either component pricing or pure bundling.

APPENDIX A

We noted in the text that BSP may be more profitable than PB if:

 (i) Willingness to pay for the bundle of all K products is heterogeneous across consumers; and

 (ii) Consumers (or consumer types) who have the highest willingness to pay for a bundle of size m are not necessarily the same as those with the highest willingness to pay for a bundle of size $n > m$.

We can establish a more formal condition by using an approach similar to McAfee et al. They note that MB nests CP as a special case, and derive a condition on the joint distribution of tastes such that a local deviation from the CP prices yields an increase in profits. In our case, we know that BSP nests PB as a special case, and we can ask when a local deviation from the PB price will be profitable.

Since BSP allows consumers to pick their own bundles, any purchased bundle of m products will consist of the m products for which the consumer's valuations were highest. Let r_{im} denote consumer i's mth -highest valuation (i.e., the mth order statistic). Then consumer i's willingness to pay for a bundle of size m is just $y_{im} = \sum_{k=1}^{m} r_{ik}$; i.e., the sum of the first m order statistics. Using this notation, we can write a sufficient condition for BSP to yield higher expected profits than PB in terms of the joint distribution of $y_{i,K-1}$ and r_{iK}:

PROPOSITION: *Suppose a firm sells K products for which marginal costs are identical and equal to c, and let g denote the joint distribution of $y_{i,K-1}$ (a consumer's willingness to pay for a bundle of any $K - 1$ products) and r_{iK} (the willingness to pay for the least preferred product). If p^* is the optimal PB price, then BSP is more profitable than PB if there exists a Δ such that*

$$(i)\ 0 < \Delta < c,$$

$$(ii)\ \int_0^{\Delta} \int_{p^*-r}^{\infty} g(y, r)\,dy\,dr > 0.$$

To prove this, consider starting with BSP prices equal to the optimal PB price, $p_1 = p_2 = \ \ldots \ = p_K = p^*$, and then reducing the price of bundles with $K - 1$ or fewer products to $\tilde{p}_{K-1} = p^* - \Delta$. The expected profits under these prices are

$$\tilde{\pi}(\Delta) = (p^* - Kc) \int_{\Delta}^{\infty} \int_{p^*-r}^{\infty} g(y, r)\,dy\,dr$$

$$+ (p^* - \Delta - (K - 1)c) \int_0^{\Delta} \int_{p^*-\Delta}^{\infty} g(y, r)\,dy\,dr.$$

The difference from the PB profits is then

$$\tilde{\pi}(\Delta) - \tilde{\pi}(0) = (c - \Delta)\int_0^\Delta \int_{p^*-r}^\infty g(y, r)\,dy\,dr$$

$$+ (p^* - \Delta - (K - 1)c)\int_0^\Delta \int_{p^*-\Delta}^{p^*-r} g(y, r)\,dy\,dr.$$

The profit difference is the sum of two effects. The first term reflects the incremental cost savings from not selling to consumers who value the Kth good below its cost but who buy the full bundle under PB pricing. The second term is the incremental profit from inducing some consumers who do not purchase under PB pricing to buy the $K - 1$ bundle. Conditions (i) and (ii) of the proposition simply guarantee that both of these effects are positive.

Note that when marginal cost is zero, condition (i) will not be met. However, this does not mean BSP cannot be more profitable than PB when marginal cost is zero: the proposition establishes a sufficient but not necessary condition for BSP profits to be higher than PB profits. So even if this kind of local change is not profitable, there may still be other (nonlocal) changes that are. However, the proposition does suggest that positive marginal costs make it more likely that BSP beats PB.

APPENDIX B

In this Appendix we simply report the optimal prices and profits for the two-good model described in Section IIIB. Consumers' valuations for the two goods are independent uniform random variables on $[0,1]$ and $[0,\theta]$, respectively. (Assume $\theta \geq 1$.) Marginal cost is 0.

Scheme	Optimal prices	Optimal profits
CP	$p_1^* = \frac{\theta}{2},\ p_2^* = \frac{1}{2}$	$\pi^* = \frac{(1+\theta)}{4}$
PB	$p^* = \begin{cases} \sqrt{\frac{2\theta}{3}} & \text{if } \theta \leq 3/2 \\ \frac{1}{4} + \frac{\theta}{2} & \text{if } \theta > 3/2 \end{cases}$	$\pi^* = \begin{cases} \left(\frac{2}{3\theta}\right)^{3/2} & \text{if } \theta \leq 3/2 \\ \frac{1}{8\theta}\left(\theta + \frac{1}{2}\right)^2 & \text{if } \theta > 3/2 \end{cases}$
BSP	If $\theta \leq 1.756739614$: $p_2^* = \frac{(1+\theta)}{3}$ $p_2^* = \frac{1}{3}\left(2 + 2\theta - \sqrt{2\theta^2 - 2\theta + 2}\right)$ Otherwise: $p_1^* = p_2^* = \frac{1}{4} + \frac{\theta}{2}$	If $\theta \leq 1.756739614$: $\pi^* = \frac{(2\theta^2 - 2\theta + 2)^{3/2} - 3\theta^3 + 9\theta^2 + 9\theta - 3}{27\theta}$ Otherwise: $\pi^* = \frac{1}{8\theta}\left(\theta + \frac{1}{2}\right)^2$
MB	If $\theta \leq 2$: $p_1^* = \frac{2\theta}{3},\ p_2^* = \frac{2}{3}$ $p_{12}^* = \frac{2}{3} + \frac{2\theta}{3} - \frac{\sqrt{2\theta}}{3}$ If $\theta > 2$: $p_1^* = \frac{\theta}{2} + \frac{1}{3},\ p_2^* = \frac{2}{3}$ $p_{12}^* = \frac{\theta}{2} + \frac{1}{3}$	If $\theta \leq 2$: $\pi^* = \frac{2}{9}\left(1 + \theta + \frac{1}{3}\sqrt{2\theta}\right)$ If $\theta > 2$: $\pi^* = \frac{27\theta^2 + 36\theta - 4}{108\theta}$

REFERENCES

Adams, William James, and Janet L. Yellen. 1976. "Commodity Bundling and the Burden of Monopoly." *Quarterly Journal of Economics*, 90(3): 475–98.

Armstrong, Mark. 1999. "Price Discrimination by a Many-Product Firm." *Review of Economic Studies*, 66(1): 151–68.

Bakos, Yannis, and Erik Brynjolfsson. 1999. "Bundling Information Goods: Pricing, Profits, and Efficiency." *Management Science,* 45(12): 1613–30.

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

Byzalov, Dmitri. 2008. "Unbundling Cable Television: An Empirical Investigation." http://www.gsb.stanford.edu/facseminars/events/marketing/documents/mktg_09_09_byzalov.pdf.

Chu, Chenghuan Sean. 2008. "The Effect of Satellite Entry on Product Quality for Cable Television." Federal Reserve Board Divisions of Research & Statistics and Monetary Affairs Finance and Economics Discussion Paper 2008–12.

Crawford, Gregory S. 2008. "The Discriminatory Incentives to Bundle in the Cable Television Industry." *Quantitative Marketing and Economics*, 6(1): 41–78.

Crawford, Gregory S., and Joseph Cullen. 2007. "Bundling, Product Choice, and Efficiency: Should Cable Television Networks Be Offered a La Carte?" *Information Economics and Policy*, 19(3–4): 379–404.

Crawford, Gregory S., and Ali Yurukoglu. 2009. "The Welfare Effects of Bundling in Multi-Channel Television Markets." http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/crawford/research/bundlingpricing.pdf.

Dubé, Jean-Pierre. 2004. "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks." *Marketing Science,* 23(1): 66–81.

Fang, Hanming, and Peter Norman. 2006. "To Bundle or Not to Bundle." *RAND Journal of Economics*, 37(4): 946–63.

Geng, Xianjun, Maxwell B. Stinchcombe, and Andrew B. Whinston. 2005. "Bundling Information Goods of Decreasing Value." *Management Science,* 51(4): 662–67.

Gill, Philip E., Walter Murray, and Michael A. Saunders. 2002. "SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization." *Society for Industrial and Applied Mathematics Journal of Optimization,* 12(4): 979–1006.

Hajivassiliou, Vassilis A., Daniel L. McFadden, and Paul Ruud. 1996. "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results." *Journal of Econometrics*, 72(1–2): 85–134.

Hendel, Igal. 1999. "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns." *Review of Economic Studies*, 66(2): 423–46.

Ho, Justin, Katherine Ho, and Julie Holland Mortimer. 2008. "Welfare Effects of Full-Line Forcing Contracts in the Video Rental Industry." http://www.columbia.edu/~kh2214/papers/hohomortimer_121008.pdf.

Jedidi, Kamel, Sharan Jagpal, and Puneet Manchanda. 2003. "Measuring Heterogeneous Reservation Prices for Product Bundles." *Marketing Science,* 22(1): 107–30.

Leslie, Phillip. 2004. "Price Discrimination in Broadway Theater." *RAND Journal of Economics*, 35(3): 520–41.

Leslie, Phillip, and Alan Sorensen. 2009. "The Welfare Effects of Ticket Resale." http://www.stanford.edu/~pleslie/resale.pdf.

Malueg, David A., and Christopher M. Snyder. 2006. "Bounding the Relative Profitability of Price Discrimination." *International Journal of Industrial Organization,* 24(5): 995–1011.

Manelli, Alejandro M., and Daniel R. Vincent. 2006. "Bundling as an Optimal Selling Mechanism for a Multiple-Good Monopolist." *Journal of Economic Theory*, 127(1): 1–35.

McAfee, R. Preston. 2002. "Coarse Matching." *Econometrica*, 70(5): 2025–34.

McAfee, R. Preston, John McMillan, and Michael D. Whinston. 1989. "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values." *Quarterly Journal of Economics*, 104(2): 371–83.

McFadden, Daniel L. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica*, 57(5): 995–1026.

Neeman, Zvika. 2003. "The Effectiveness of English Auctions." *Games and Economic Behavior*, 43(2): 214–38.

Pakes, Ariel, and David Pollard. 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica*, 57(5): 1027–57.

Rogerson, William P. 2003. "Simple Menus of Contracts in Cost-Based Procurement and Regulation." *American Economic Review*, 93(3): 919–26.

Salinger, Michael A. 1995. "A Graphical Analysis of Bundling." *The Journal of Business*, 68(1): 85–98.

Schmalensee, Richard. 1984. "Gaussian Demand and Commodity Bundling." *Journal of Business*, 57(1): S211–30.

Shiller, Ben, and Joel Waldfogel. 2009. "Music for a Song: An Empirical Look at Uniform Song Pricing and Its Alternatives." http://bpp.wharton.upenn.edu/waldfogj/pdfs/Shiller_Waldfogel_submission.pdf.

Spence, A. Michael. 1980. "Multi-Product Quantity-Dependent Prices and Profitability Constraints." *Review of Economic Studies*, 47(5): 821–41.

Stigler, George J. 1963. "United States v. Loew's Inc.: A Note on Block-Booking." *Supreme Court Review,* 1963(1963): 152–57.

Venkatesh, R., and Vijay Mahajan. 1993. "A Probabilistic Approach to Pricing a Bundle of Products or Services." *Journal of Marketing Research,* 30(4): 494–508.

Wilson, Robert. 1993. *Nonlinear Pricing*. New York: Oxford University Press.

This article has been cited by:

1. Yongmin Chen, Michael H. Riordan. 2013. PROFITABILITY OF PRODUCT BUNDLING*. *International Economic Review* **54**:1, 35-57. [CrossRef]

2. Justin Ho, Katherine Ho, Julie Holland Mortimer. 2012. Analyzing the Welfare Impacts of Full-line Forcing Contracts. *The Journal of Industrial Economics* **60**:3, 468-498. [CrossRef]

3. Anja Lambrecht, Katja Seim, Naufel Vilcassim, Amar Cheema, Yuxin Chen, Gregory S. Crawford, Kartik Hosanagar, Raghuram Iyengar, Oded Koenigsberg, Robin Lee, Eugenio J. Miravete, Ozge Sahin. 2012. Price discrimination in service industries. *Marketing Letters* **23**:2, 423-438. [CrossRef]

4. Katherine Ho,, Justin Ho,, Julie Holland Mortimer. 2012. The Use of Full-Line Forcing Contracts in the Video Rental Industry. *American Economic Review* **102**:2, 686-719. [Abstract] [View PDF article] [PDF with links]