NBER WORKING PAPER SERIES

NEARLY OPTIMAL PRICING FOR MULTIPRODUCT FIRMS

Chenghuan Sean Chu
Phillip Leslie
Alan Sorensen

Working Paper 13916
http://www.nber.org/papers/w13916

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2008

We are grateful to Amitay Alter for research assistance, and to TheatreWorks for providing the data. Thanks also to Lanier Benkard, Ken Corts, Sanjog Misra, Peter Reiss, Garth Saloner, Andy Skrzypacz and Mike Whinston for helpful discussions. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Board of Governors or the staff of the Federal Reserve System. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER WORKING PAPER SERIES

NEARLY OPTIMAL PRICING FOR MULTIPRODUCT FIRMS

Chenghuan Sean Chu
Phillip Leslie
Alan Sorensen

Working Paper 13916
http://www.nber.org/papers/w13916

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2008

We are grateful to Amitay Alter for research assistance, and to TheatreWorks for providing the data. Thanks also to Lanier Benkard, Ken Corts, Sanjog Misra, Peter Reiss, Garth Saloner, Andy Skrzypacz and Mike Whinston for helpful discussions. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Board of Governors or the staff of the Federal Reserve System. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Nearly Optimal Pricing for Multiproduct Firms
Chenghuan Sean Chu, Phillip Leslie, and Alan Sorensen
NBER Working Paper No. 13916
April 2008
JEL No. D4,L0,L11

ABSTRACT

In principle, a multiproduct firm can set separate prices for all possible bundled combinations of its products (i.e., "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. In this study we show that simple pricing strategies are often nearly optimal -- i.e., with surprisingly few prices a firm can obtain 99% of the profit that would be earned by mixed bundling. 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. These findings are based on an array of numerical experiments covering a broad range of demand and cost scenarios, as well as an empirical analysis of the pricing problem for an 8-product firm (a theater company).

Chenghuan Sean Chu
Federal Reserve Board of Governors
Sean.Chu@frb.gov

Phillip Leslie
Stanford University
Graduate School of Business
518 Memorial Way
Stanford, CA 94305
and NBER
pleslie@stanford.edu

Alan Sorensen
Graduate School of Business
Stanford University
518 Memorial Way
Stanford, CA 94305-5015
and NBER
asorensen@stanford.edu

1 Introduction

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 10 products, there are over a thousand 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 99% of the profit that 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 (BSP)—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.

BSP involves setting different prices for different sized bundles. For a firm with 3 goods, BSP sets one price for the purchase of any single good, a second price for the purchase of any 2 goods, and a third price for purchasing all 3. 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]

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 10 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

[1]Component pricing is sometimes also referred to as "independent good pricing".
[2]See McAfee, McMillan and Whinston (1989).
[3]See Stigler (1963) and Adams and Yellen (1976).

as a single package may be more profitable than offering them individually (or perhaps not). Our findings offer a new suggestion: BSP requires only 10 prices and attains 99% of the profit from 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] However, it may seem that CP would be more profitable when there is a high degree of demand asymmetry across products. In fact, BSP is also able to extract surplus from individuals with high demand for one product and not others—BSP does so 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 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.

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. 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.[6] An obvious limitation to this approach 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. This allows us to demonstrate that our findings apply to an empirically relevant model.

The empirical analysis is based on a theater company that produces a season of 8 plays. It

[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 Bakos and Brynjolffson (1999).

[6] We are not the first to rely on numerical methods to analyze bundling problems. See also Schmalensee (1984) and Fang and Norman (2006).

is an interesting setting in which to compare the profitability of different pricing schemes. On the one hand, the plays differ in overall popularity, suggesting that component pricing may be important for profits. On the other hand, many consumers attend multiple plays, suggesting that some form of bundling may also be profitable. With 8 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 to 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 covariances 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% less profitable than CP in this case. Nevertheless, we find that BSP is 0.9% more profitable than CP, and BSP attains 98.5% of the MB profits.[7]

While our focus is on bundling, approximating complex strategies using simpler alternatives is also an important issue in the theory of contracts. 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] See also McAfee (2002) and Wilson (1993) for similar findings in the context of nonlinear pricing.

The remainder of the paper is organized as follows. In Section 2 we summarize the relevant prior literature. Section 3 describes the basic intuition underlying the various pricing strategies, and presents the results from an extensive set of numerical experiments. The empirical example is presented in Section 4. Section 5 concludes.

[7] If we set the estimated covariances in the demand system to zero, holding all other estimated parameters fixed, BSP is 20.5% more profitable than CP.

[8] In the specific case of uniform types and quadratic effort costs, he shows that exactly 75% 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.

2 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 2 products, and consumers vary in their willingness-to-pay for each product, then Stigler (1963) shows by example that selling these 2 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 willingness-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' valuations is non-negative.[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, 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).

[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 non-trivial ways, as noted by Salinger (1995). In all of the analysis in our study we allow for free disposal.

[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 2 that is higher than the sum of component prices. We limit our analysis to the no-monitoring case. See also Manelli and Vincent (2006).

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

Armstrong (1999) provides a more general but similar result to 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 5 products, say.

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 UP.

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 there exist pricing schemes that involve few enough prices to be feasible, and that tend to yield profits that are close to the MB level?

3 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.[14] 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

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

alternatives. Similarly, CP nests UP as a special case, and BSP nests both UP and PB as special cases. CP and BSP are non-nested alternatives.

CP and BSP are of particular interest because the number of prices equals the number of products, which is a reasonable benchmark for practical pricing strategies. However, there are many other potential pricing strategies that also involve K prices that are nested subsets of MB. The problem in these cases is that it is *ex ante* unclear which subset of K prices to choose. There are also strategies (with more than K prices) that nest both CP and BSP.

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 that involves N prices, for each value of N from 1 to $2^K - 1$, given a particular model of demand and costs.[15] Such an upper-bound is 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. 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.[16] Second, in any example with more than a few goods, there is a large number of possible pricing structures that must be evaluated, which is a non-trivial computational challenge. For these reasons, a complete answer to this question is beyond the scope of this study.

However, it should be noted that one of our main findings is that BSP usually attains around 99% 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% 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).[17] For the 2006 season the Los Angeles Dodgers offered several bundles of specific games, a discount for

[15] Malueg and 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.

[16] 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.

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

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 the Giants 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 5-play season. In 2006-07 they offered a discount for the 5-play bundle at a variety of prices that varied by time-of-week, and equal prices for individual shows (also varying by time-of-week). In 2006-07 the San Francisco Opera had a 10-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 3 or more products.

3.1 Examples with Two Goods and Two Consumer Types

In order to clarify the basic intuition that underlies the various pricing schemes, in this section we present select examples with two goods and two consumer types in which the optimal prices under the various schemes are simple to determine. The examples illustrate how each of CP, PB, and BSP may attain the highest profits in different settings. We 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 or substitutes); and (iii) there is no resale.

The seminal papers on bundling pointed out that PB can be more profitable than CP because it may reduce heterogeneity in consumers' willingness-to-pay. Consider the following example. There are two consumers (A and B) and two products (1 and 2) with zero marginal costs. Each consumer has valuations v_1 and v_2 for the two products. Valuations are given by:

	v_1	v_2
A	4	1
B	1	4

In this case the optimal CP prices are 4 and 4, and the CP profit is 8. With PB both consumers value the bundle at 5, and so the PB price is 5, extracting the full surplus of 10. This is a textbook example of why bundling can increase profits even though there are no complementarities in demand or costs.

When is CP more profitable than PB? Intuition suggests that if the optimal CP price would be much higher for one product than the other, CP is likely to be better. But asymmetry of this sort isn't enough. Consider the example:

	v_1	v_2
A	4	1
B	4	1

Demand for good 1 is higher than for good 2, and CP accommodates this by charging a price that is 4 times higher for good 1. The CP profit is 10. But both consumers also equally value the bundle at 5, allowing PB to also obtain a profit of 10. CP does no better than PB, despite the substantial demand asymmetry. Note also that the above valuations provide an example of vertically differentiated products—both consumers agree that good 1 is preferred to good 2. It is interesting that PB may be equally profitable to CP in such a case.

For CP to significantly outperform PB, the valuations must exhibit a kind of within-product asymmetry. In particular, a large fraction of the extractable surplus must be concentrated on one product and one consumer type. For example:

	v_1	v_2
A	1	2
B	5	2

In this case CP charges 5 and 2, and extracts 9. PB sets a price of 7, sells only to type B and extracts 7. Intuitively, CP is able to charge a high price for product 1 in order to extract the large amount of surplus attributable to type B, but this does not rule out selling good 2 to both types. PB, on the other hand, by extracting as much surplus as possible from type B, ends up abandoning type A altogether.

So far we have shown how PB may be more profitable than CP, and how CP may be more profitable than PB. What about BSP? In all of the above examples BSP is equivalent to PB. This is trivially true in the first two examples, because PB extracts the full surplus.[18] In the last example, it is tempting to set BSP prices of 2 (for one good) and 7 (for two goods), but in that case type B would choose only good 1 rather than the bundle of two, so BSP can do no better than PB.[19]

[18]Note that PB can never yield strictly higher profit than BSP because PB is a constrained version of BSP.

[19]Put differently: in order to induce B to buy the bundle, the BSP price for the bundle of one has to be

Clearly BSP cannot do worse than PB. But under what circumstances will BSP be strictly more profitable than PB? Consider the following example:

	v_1	v_2
A	4	0
B	3	3

The optimal CP prices are 3 and 3 yielding a profit of 9. The optimal PB price is 4, which yields a profit of 8. BSP charges 4 for the purchase of any single good, and 6 for the purchase of both. Type A buys product 1, type B buys the bundle of both, and profits are 10. Intuitively, the reason BSP is more profitable is that the consumer with the highest valuation for a bundle of one is different from the consumer with the highest valuation for a bundle of two. BSP is able to extract more surplus by setting prices that separate the two consumers, whereas PB is forced to pool the two types by lowering the price of the bundle of two.

Loosely speaking, we expect BSP to outperform PB when (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 $n > m$. In Appendix A we provide a formal condition that is sufficient for BSP to be strictly more profitable than PB.

Note also, in this example both consumers weakly prefer good 1 to good 2 (vertically differentiated products). This is analogous to a baseball team with one specific game that all consumers value more than any other game. Economists sometimes cite the overwhelming popularity of certain baseball games, such as games between traditional rivals like the New York Yankees and the Boston Red Sox, as evidence that CP would be much more profitable than UP. It is interesting that in such a setting BSP may be even more profitable than CP.

3.2 An Example with Two Goods and a Continuum of Consumer Types

The goal of any pricing strategy is to extract as much surplus as possible from consumers. If consumers are heterogeneous, then a strategy that can separate consumers according to willingness-to-pay will extract more surplus than a strategy that cannot. CP separates consumers in a straightforward way: consumers with a high valuation for a given good can be

high—but doing this means that A doesn't buy anything. The best BSP can do is pool the two types together by charging the PB price.

separated from consumers with a low valuation. In the two type example, this can result in one type buying a single good and the other type buying both goods. The same pattern can arise under BSP, but the mechanism is quite different. For BSP to obtain separation of the two types, the type with the highest valuation for a single-good bundle must differ from the type with the highest valuation for the two-good bundle. If not, then BSP is equivalent to PB and there is pooling. Of course, this condition is inconsequential for whether there is separation under CP, which illustrates that CP and BSP are very different screening devices.

With a continuum of consumer types, the difference in screening becomes more complex. Consider the following example. Assume there are two goods, both with zero marginal cost. Consumers' valuations for good 1 are uniformly distributed between zero 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).[20] If $\theta = 1.7$ then the optimal CP prices are .85 and .5 for goods 1 and 2, respectively. Under BSP the optimal price for a single-good bundle is .9, and the price for the bundle of both goods is 1.1. In this example BSP is 5.6% more profitable than CP (even though the optimal CP prices vary by 70% 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 .67, and the price for the bundle is 1.18. Unsurprisingly, the price for a single-good bundle under BSP (.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% of the total profit under MB comes from consumers who buy one good, with the remaining 85% 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 being more important 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.

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 .85 purchase good 1, and consumers above .5 purchase good 2 (with consumers in region A purchasing both). Under BSP consumers in the two regions labelled 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

[20] 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.

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 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 (.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 on getting consumers to purchase multiple goods than they would have under CP. Relative to CP, BSP raises the price for single-good buyers, and lowers the price for multi-good buyers. It is profitable to do this in this example, but there is 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 that 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 would increase the relative profitability of BSP. Note the downward trend of the shaded regions in which BSP extracts greater surplus than CP—negative correlation would tend to increase the fraction of consumers in these regions.[21] 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

[21] 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.

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 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 that 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 4, 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 non-rectangular and non-contiguous (i.e. region C). Adding more goods or incorporating non-zero 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.

3.3 Numerical Analysis with Continuous Types and More than Two Goods

Although the two-good examples described above are illustrative, our objective is to analyze multiproduct pricing strategies in a more general context, and in particular to allow for more than two products. As we have argued, when the number of products increases MB quickly becomes highly complex. So it is important to understand which subset of prices (if any) can capture a large fraction of the profit that MB would obtain. The above examples illustrate how any of CP, PB or BSP may be the most profitable in any given circumstance.

The results in this section are based on a broad range of computational experiments in which we solve for the optimal prices and profits for 5 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.[22] 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. 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.[23]

We vary the number of goods in the experiments from 2 to 5. Each experiment is performed under four different assumptions regarding costs: (i) all products have zero marginal cost; (ii) all products have positive and equal marginal cost; (iii) all products have positive but differing marginal cost (we set marginal costs equal to half of each product's mean valuation); and (iv) marginal costs are zero but there is a binding capacity constraint.[24]

Table 2 describes 13 alternative assumptions on the distribution of consumers' valuations (F) that we consider in our experiments. Note that we include distributions with non-zero covariances in product valuations across products. This is important since correlation in tastes is a key determinant of the profitability of bundling, as the prior literature has noted. Exponential, logit, lognormal, and normal distributions are all commonly used in empirical studies of demand. The uniform distribution is often convenient in theoretical studies of demand and is also occasionally used in empirical work.

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. In each case the boundaries were chosen so that the range of optimal prices is roughly similar across cases to help with comparability. It is also important that our experiments include cases with a high degree of demand asymmetry, which can favor CP. Hence, we choose grid boundaries that allow optimal CP prices to vary by up to a factor of 10.[25]

[22]As in the two-type examples in the prior subsection, by assuming additive preferences we are ruling out consumption complementarities as a motivation for bundling.

[23]Schmalensee (1984) does not allow free disposal. Either assumption may be correct depending on the particular products.

[24]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 .9 times the demand for the most popular product 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.

[25]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 than 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 game would have have an optimal component price that is several times greater (though probably not more than 10 times greater) than the least popular game. See also our empirical example in the next section. Nevertheless, it must be noted that some of our results may not generalize to settings where demand differs more dramatically across products.

The granularity of each grid of parameter values varies with the number of products, so that we analyze approximately 220 parameter combinations for each class of distribution regardless of the number of products. We consider 13 parametric families, 4 marginal cost assumptions, variation in the number of products from 2 to 5, and about 220 parameter combinations in each case—leading us to compute 5 sets of prices and profits in over 45,000 different examples. Numerical methods are used to find the optimal prices in each case.[26] We calculate the demands for each bundle using a kernel-smoothed frequency simulator, as discussed in Hajivassiliou, McFadden, and Ruud (1996), using 10,000 simulated consumers and a logistic kernel with smoothing parameter 0.02.

Before summarizing the outcomes of these experiments, 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 over-sampling the relevant (i.e., empirically plausible) combinations of parameters. So, for example, when we describe average outcomes, these should certainly *not* be interpreted as outcomes that would be expected in an empirical sense—they should be interpreted narrowly as the average of the experiments we performed.

3.4 Results from Numerical Analysis

Figure 2 provides a summary of the numerical experiments for 3 different assumptions about costs.[27] The figure shows box-plots depicting various percentiles of the distribution of profits

[26] Specifically, we use SNOPT, a sequential quadratic programming algorithm developed by Gill *et al* (2002) for solving nonlinear constrained optimization problems. For BSP and MB, we also check to make sure the computed optimal prices are robust to alternative start values.

[27] We exclude the experiments for positive and equal marginal costs because they add no further insight, but we do include these results in the tables we discuss below.

under each pricing strategy relative to BSP. To construct a given box-plot we pool experiments across distributions of consumers' valuations and for $K = 2, ..., 5$.[28] Hence, while the figures reveal the range of outcomes, they hide the differences across distributions and across K.[29] 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.[30]

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 to be taken from Figure 2:

1. BSP tends to be more profitable than CP. Based on the 46,344 experiments we performed (across different cost assumptions and across different taste distributions), we find that BSP is more profitable than CP 91% of the time. Furthermore, BSP obtains 13% higher profit than CP, on average.

2. BSP tends to obtain profits that are within 1% of the profits from MB. Specifically, the profit from BSP is within 1% of MB in 60% of the 46,344 experiments we performed. And on average, we find that BSP yields 98% of the MB profits.

Figure 2 also shows that varying assumptions about costs has an impact on the relative profits of the different pricing strategies, but the effect is quite small. Under the assumption of zero marginal costs, BSP is more profitable than CP in 97% of the experiments, and BSP is within 1% of the MB profits in 75% of the experiments. In comparison, under the assumption of positive and unequal marginal costs, BSP is more profitable than CP in 87% of the experiments, and BSP is within 1% of the MB profits in 34% of the experiments (although even here BSP attains 97% of the MB profits, on average). This is to be expected since the prior literature has explained that increases in marginal costs make "exclusion" (i.e., preventing consumers from purchasing goods they value below marginal cost) relatively more important.

In the introduction we noted that prior research shows that PB can be more profitable than CP for finite K. We find that PB attains higher profit than CP in 61 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 53, 61, 64, and 66 percent of the experiments, respectively. Fang and Norman (2006) also find this pattern in their numerical experiments.

[28] Note that we pool across parameter combinations for a given parametric family as well as pooling across parametric families (in addition to pooling across K).

[29] Those differences are shown in Tables 3 to 6, discussed below, and in even more detail in Appendix C.

[30] 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.

In Tables 3 through 6 we summarize the results at a less aggregate level, showing how the performance of each pricing strategy varies according to the parametric family for the joint distribution of consumers valuations (under each cost scenario). We report the 1st, 50th and 99th percentiles of the distribution of profits of each pricing strategy relative to the profit from BSP, for each parametric family of consumers'valuations. Hence, we pool together the results from experiments with differing numbers of products ($K = 2, ..., 5$).[31] In Appendix C we provide more detailed summary statistics. To conserve space, we omit 5 of the parametric families described in Table 2 from Tables 3 through 6, because they make no qualitative difference to any of the findings.[32] The detailed results for these distributions are, however, included in Appendix C.

The main point to take from Tables 3 through 6 is that the choice of parametric family may not be innocuous in terms of the profitability of different pricing strategies. For instance, for the logit distribution (which is one of the most commonly used in empirical research) BSP is always more profitable than CP regardless of the level of marginal costs.[33] The same is true for log-normal distributions.

The role of a given parametric family may also vary depending 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 exponentially distributed. 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) is sufficiently unrestrictive in the sense that either CP or BSP may be the most profitable under any assumption on costs. In the empirical example we analyze below, we assume normally distributed tastes.

3.5 Discussion of Numerical Analysis

The numerical experiments demonstrate that BSP is sometimes more profitable than CP, and that CP is sometimes more profitable than BSP (although the former is much more common in our experiments). A key theme of this study is that BSP performs well even in the presence of a high degree of demand asymmetry, as we explained in the two-good examples with analytic

[31] To be more precise, for each combination of parameters of the taste distribution we calculate the ratio of profits under pricing strategy X to profits under BSP. The table reports various percentiles of this ratio across parameter combinations and across $K = 2, ..., 5$. There are around 900 experiments in each distribution.

[32] We leave out: Lognormal(-), Lognormal(+), Normal(+/-), Normal(v-), and Normal(v+).

[33] Although the presence of capacity constraints can favor CP over BSP with logit demand.

solutions. However, intuitively we still expect that increasing demand asymmetry favors CP over BSP.

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 vs. BSP. Even if we look at the top 10% of experiments in terms of demand asymmetry (as measured by the price ratio), we find that BSP is still more profitable than CP in over 80% of these experiments. In other words, a high degree of asymmetry does *not* imply that CP will be more profitable than BSP.[34]

To better understand why BSP tends to obtain higher profits than CP in our numerical experiments, even when demand is highly asymmetric, in Table 7 we compare prices and market shares under CP, BSP and MB. The table documents how close the CP prices and BSP prices are to the MB prices, as well as the closeness of the market shares. For each possible bundle of a given size we compute absolute 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 29.5%. In contrast, BSP prices in the same experiments tend to differ from MB prices by 65.7%. Hence, Table 7 reveals that CP prices for small-sized bundles (for any given K) tend to be closer to the MB prices than BSP does. But 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.

The fact that prices for large-sized bundles under BSP tend to be close to the MB prices for the same bundles stems from two sources. Consider an example in which there are 5 goods ($K = 5$), and consider the prices for the various bundles containing 4 of these 5 goods (there are 5 such bundles). Under BSP these bundles are equally priced, while under MB there may be 5 different prices for these bundles. The results in Table 7 indicate that: (i) the average price of these 5 bundles under MB is close to the uniform price under BSP; and (ii) there is not much variation in prices across these 5 bundles under MB. The second of these features is an interesting consequence of heterogeneity-reduction. That is to say, as bundle-size increases, the demand for alternative bundles of the same size becomes similar. Hence, different bundles of the same size do not need to be priced very differently if the bundles are large. This is why BSP prices tend to be an especially good approximation of MB prices for large-sized bundles.

[34] It is possible that some other measure of demand asymmetry is a better predictor of relative profits. We have explored several alternatives, but we have not found any single summary statistic (or collection of summary statistics) based on demand asymmetry that serves as a good predictor of relative profits.

Consider also the market shares shown in Table 7. 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 $K = 3$, CP sells a single good to 38% of consumers, while BSP and MB sell a single good to only 12% and 14%, respectively. Meanwhile, BSP and MB sell the full bundle to 29% and 27% of consumers, respectively, and CP sells the full bundle to only 8% 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.

In summary, there are two reasons why BSP tends to perform so well compared to CP even in the presence of a high degree of demand asymmetry. First, most of the profit under BSP derives from consumers that purchase multiple goods, and there is little benefit from varying prices across larger-sized bundles with equal number of products. Second, it needs to be a rather particular form of demand asymmetry for BSP to be less profitable than CP (see Section 3.1). For example, suppose a firm sells 3 goods that are vertically differentiated and the optimal prices under CP are 1,000, 10, and 1. This is an extremely high degree of demand asymmetry, yet BSP can also perform well in this case, by setting a price for any one good of around 1,000, a price for any two of around 1,010 and a price for all 3 of around 1,011.

Table 8 shows the consequences for social surplus. 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 also reduce the dead weight loss by significant amounts, relative to CP. Interestingly, the table also indicates that BSP and MB tend to 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.

As noted in the beginning of this section, aside from BSP there are many other potential pricing schemes that require the firm to set K prices for different bundles of goods. It would be interesting to compare the profits from BSP with the distribution of profits for all other K-price schemes, for a given demand specification. It is possible that another K-price strategy would be more profitable than BSP in any given demand specification. Whether there exists a particular K-price strategy that attains higher profits than BSP across all possible specifications is questionable. We make no attempt to perform this analysis here because it imposes a significant computational burden. For a single demand specification with 5 goods ($K = 5$) there are nearly 170,000 different pricing schemes that involve 5 prices. Moreover, it would be important to implement this analysis for many different demand specifications, numbers of prices, and numbers of goods. We therefore leave this to future research.

Lastly, a potential concern is that BSP is such a close approximation to MB because prices simply don't 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. In unreported experiments we found this tends to yield around 15% lower profit than if the firm had correctly assumed positive correlations.[35] This provides a degree of assurance that profits are indeed sensitive to prices in our experiments.

4 Estimation of Joint Distribution of Consumers' Valuations

An obvious limitation of the numerical experiments in Section 3 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 8-play season. Given these estimates, we compute the profitability of each pricing strategy, allowing us to demonstrate that our findings apply to an empirically relevant model.

In this section we address the problem of how to estimate the joint distribution of consumers' valuations from available data. It is important that such an approach allow for non-zero covariances in tastes, because covariance is a major determinant of the relative profits from different bundling-type schemes. This rules out one commonly made assumption in demand estimation: that unobserved tastes have a logit distribution. Of course, we are not the first to estimate covariances in a demand system: recent examples include Bajari and Benkard (2005) and Hartmann and Viard (2007); see also Allenby and Rossi (1999) for a review of the earlier literature on flexible estimation of demand heterogeneity. It is also important that our approach allow for consumers to purchase multiple products. This is somewhat non-standard in demand estimation based on discrete choice models where it is commonly assumed that consumers choose at most one product. However, at least a couple of prior papers have relaxed this assumption to allow for multiple purchases in discrete choice settings, such as Dubé (2004) and Hendel (1999).

Several features make our particular empirical example an appealing context in which to

[35]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.

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 are confident there is no significant resale activity—these are plays produced by a small theater company, not rock concerts or professional sports.[36] 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.

A by-product of the analysis is that we measure the impact of each pricing strategy on consumer welfare. This is interesting because bundling, like price discrimination more generally, has ambiguous effects on consumer welfare relative to uniform pricing.[37] To the best of our knowledge, there is one prior empirical analysis of bundling. Crawford (2006) tests the hypothesis that consumers' 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. Based on a calibrated demand model, Crawford argues that adding a top-15 cable channel to a bundle and re-optimizing prices leads to 5.5% lower consumer surplus, and 6.0% higher profit.[38]

4.1 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 8 different plays or musicals. Table 9 provides summary information for each of the 8 plays. A total of 69,207 tickets were sold to the 8 plays.

Consumers could purchase tickets to individual plays at a uniform price, but most of the tickets (80%) were purchased as part of a subscription. *TheatreWorks* offered 3 subscription packages: (i) the full 8-play season; (ii) any combination of 5 plays; or (iii) a pre-specified bundle of 3 plays.[39] 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.

[36] See Leslie and Sorensen (2007) for an empirical analysis of ticket resale.

[37] 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.

[38] Two studies in the marketing literature use survey response data to estimate demand and compare profits from UP, PB and MB: Venkatesh and Mahajan (1993) and Jedidi, Jagpal and Manchanda (2003).

[39] The pre-specified bundle consisted of the only 3 plays that were performed at *TheatreWorks*' secondary venue, a smaller theater in Palo Alto, CA.

Table 10 summarizes the purchase options and their average prices.[40] Subscribers, by definition, purchase tickets to multiple plays. Importantly, non-subscribers may also purchase tickets to multiple plays, and indeed we observe this in the data. However, among the non-subscribers, we can only identify multiple-play purchasers if the tickets were mailed to them (in which case we observe their name and address). This was the case for 45% of the tickets purchased by non-subscribers. The rest were purchased anonymously at the box-office, and for these purchases we have no way of knowing if the individual also purchased tickets to other plays. For the purpose of Table 10, we assume the sample of non-subscribers for whom we observe mailing information is a random subset of all non-subscribers, and extrapolate the same pattern of multi-play purchases to the entire set of non-subscribers. This helps to provide a more complete description of the sales patterns in the data. We do *not* rely on this assumption in estimation, as explained below.

As shown in Table 10, there were 5,139 subscribers to the 8-play bundle, 2,794 subscribers to a 5-play bundle, and 205 subscribers to the 3-play bundle. The popularity of the flexible 5-play subscription is a particularly important feature of the data. Observing which 5 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 5-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 information is crucial to our analysis. If we had only data on aggregate sales for each play we would be unable to identify the covariance structure of demand. Berry, Levinsohn and 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.

Table 11 summarizes the correlations within the sample of *pick-5* purchases: it reports the difference between the empirical correlations of the choices and the correlations that would be expected if tastes were independent.[41] The patterns make intuitive sense. For example, tastes for *Bat Boy*, described in the brochure as a "wacky new musical," are positively correlated with tastes for *Memphis*, described as a "rafter-rattling musical comedy." Conversely, tastes for *Bat Boy* are negatively correlated with *All My Sons*, a classic Arthur Miller drama billed as an "intense, compelling tale of love, greed, and personal responsibility."

[40] 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.

[41] Since each consumer selected 5 plays from 8, the pairwise correlations will be nonzero even if tastes are independent. The expected correlation if plays are chosen independently is -1/7.

4.2 Empirical Model

The 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 $j = 1, ..., J-1$ bundles containing combinations of the $k = 1, ..., K$ products. There is also a J^{th} option for consumers which is the outside alternative. We assume the net utility to consumer i from option j is given by

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

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 consuming particular plays together.

We allow for two classes of consumers: theater-lovers and regular consumers. In fact the data support this description, as we explain below in the subsection on identification. 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:

$$\begin{aligned} V_i &= \max\{\theta_i + \epsilon_i, 0\}, \quad \text{where} \\ \theta_i &= \begin{cases} \bar{\theta} & : \quad \text{probability } \lambda \\ 0 & : \quad \text{probability } (1-\lambda); \quad \text{and} \end{cases} \\ \epsilon_i &\sim \mathrm{N}\left(\mu, \Sigma\right). \end{aligned}$$

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. In the next subsection we discuss how the data identify λ and $\bar{\theta}$.

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 we impose the restriction that all mean terms equal zero ($\mu(k) = 0$, $\forall k$), but leave the variance-covariance matrix unconstrained.[42] In fact we also estimated the model based on the restriction that all

[42] We also normalize the variance of valuations for play (1) to equal 1: $\Sigma(1,1) = 1$.

variances equal one and the mean terms are 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. 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 8 plays implies 255 possible product combinations. This includes each individual play, the preset bundle of 3, the full bundle of all 8, 56 possible combinations of 5 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. 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.

Recall that for non-subscription 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 5 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 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 non-subscription purchases. Specifically, we only use moment conditions that are based on market shares we directly observe:

- Share of consumers who chose all 8 plays (1 moment condition)
- Shares of consumers choosing specific combinations of 5 plays (56 moment conditions)
- Share of consumers choosing the pre-set bundle of 3 plays (1 moment condition)
- Overall market shares of each play: i.e., what fraction of consumers purchased a given play as part of *any* bundle (8 moment conditions).

The last set of moment conditions utilizes information from non-subscribers without imposing any assumptions about their pattern of multi-play purchases.[43]

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.[44] 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 3, a price for choosing any 5, and a price for all 8 plays.[45] 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 8 plays. We then "fill in" the 3-play and 5-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 4 to 2, 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 8 plays, are equal to the observed prices.[46]

To compute the share of the outside option ($j = 0$) 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. We explain how the price constraint identifies the market size in the next section.

Including α, Σ, $\bar{\theta}$, 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.

[43] In a previous version of this paper, we imputed multi-play purchases among "anonymous" non-subscribers using the patterns we observe for the identifiable non-subscribers (i.e., the same approach utilized in Table 10, discussed above), and estimated the model via simulated maximum likelihood. The results are very similar to those we report below.

[44] In an earlier version of this paper we estimated the demand model via simulated maximum likelihood without any price-setting conditions. This led to predicted prices that tended to be significantly lower than the observed prices.

[45] Since capacity constraints are rarely binding in the data, we assume zero marginal costs when solving the profit-maximization problem.

[46] In practice we are within $1e - 4$ of equality in these conditions.

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.

4.3 Identification

What variation in the data serves to identify each parameter of the demand model? The variance terms, $\Sigma(k,k)$, are identified by the plays' relative overall market shares: relatively high share plays must have relatively higher variances. Note, however, that the observed ranking of market shares need not be a one-to-one mapping with the estimated play variances, because the covariance terms in Σ also have an impact on choice probabilities. For example, a given play can have a high market share either because the variance in valuations is high, or because it has a strong positive correlation with another high-variance play.

The covariance terms themselves are identified by the bundle combinations chosen by multi-play buyers, such as the *pick-5* subscribers. Pairs of plays that consumers choose to bundle relatively often will have more positive covariances. We might expect the estimated covariances to be similar to the empirical covariance matrix shown in Table 11. However, the estimated covariances are based on a model in which we control for play qualities and prices, and we utilize the complete dataset.[47] Hence, we only expect some degree of similarity. Note also, while a large fraction of consumers choose to subscribe to the full season of all 8 plays, this does not necessarily imply strong positive covariances, because other features of the model can explain this particular behavior, as we explain below.

The degree of price sensitivity, α, is identified by variation in per-play prices across bundles. Because *TheatreWorks'* pricing involves discounts for larger bundles, consumers' sensitivity to price explains why market shares for larger bundles are higher than would otherwise be the case. An additional source of pricing variation comes from the fact that one specific 3-play bundle is offered at a discount (\$36.20 per play) while all other 3-play combinations have no discount (\$40.80 per play). The taste distribution alone may explain why a specific 3-play bundle is more

[47] For example, the relatively large number of full season subscribers will encourage more positive covariances in the demand estimation.

popular than other 3-play bundles. Hence, α is identified by the extent to which demand for the discounted 3-play bundle exceeds the demand implied by the taste distribution alone. We also assume there are no complementarities in demand between these particular plays, which seems reasonable in this context. Imposing the supply-side pricing condition also helps to assure a reasonable estimate of price sensitivity.

A standard concern with demand estimation is the possibility that observed prices are correlated with unobserved demand shifters, which may bias parameter estimates. However, in the estimation we integrate over all unobserved demand components. There is no remaining error term that may be correlated with observed prices. Consider, for example, the discounted 3-play bundle. We estimate the variances and covariances of the taste distribution—i.e., we control for the qualities of these plays, and we control for the tendency of consumers to want to bundle these particular plays together. The fact that this specific bundle is offered at a discount is exogenous variation for our purposes. Stated differently, we assume there are there are no bundle-specific error terms. And even if there were, endogeneity is only a concern if bundle-specific errors vary systematically by bundle size, because *TheatreWorks'* prices are in any case only dependent on the total number of plays (with the exception of one particular three-play bundle).

How do the data identify $\bar{\theta}$ and λ? This aspect of the model is important for explaining a key feature of the data. Suppose that $\bar{\theta} = 0$ (or equivalently, $\lambda = 0$). In this case, the relationship between bundle size and market share depends on the degree of correlation in tastes for plays, but in a very particular way. If play valuations are weakly or negatively correlated, the probability of a consumer having high valuations for all 8 plays is less than the probability of having high valuations for any 5 plays, say. Hence, controlling for the effect of price, the number of 8-play subscribers would be less than the number of 5-play subscribers. Similarly, the number of 5-play subscribers would be less than the number of four-play subscribers, and so forth. On the other hand, the higher the degree of (positive) correlation, the more often we ought to observe purchases of larger bundles. But in Table 10 we see that purchases by bundle size are heavily skewed toward both individual purchases as well as purchases of all eight plays. This pattern cannot be explained by a simple joint-normal distribution, because the two most commonly purchased bundle sizes convey conflicting information about the correlation in play valuations. This is why we distinguish theater-lovers in the demand model (i.e. the reason for including λ and $\bar{\theta}$).

Clearly, the relatively high fraction of 5-play and 8-play subscribers serves to identify $\bar{\theta}$ and λ. But how are these parameters separately identified? Since the number of single-play and 8-play buyers are both greater than the number of 2, 3, 4 or 5-play buyers, λ must not be too large or too small. If λ is near to one, nearly everyone is a theater-lover, and the model would predict a low level of single-play sales. If λ is near to zero, we have the same problem described in the previous paragraph. This logic implies that λ is identified by the ratio of subscribers (i.e., 5-play and 8-play buyers) to non-subscribers (i.e., buyers of fewer than 5 plays).

Applying similar logic, if $\bar{\theta}$ is very large then all theater-lovers will choose the 8-play bundle. If $\bar{\theta}$ is near zero then we have the same problem described above: we cannot explain the bimodality of market shares in the data. Therefore, the role of $\bar{\theta}$ is to deliver an accurate prediction of the ratio of 5-play subscribers to 8-play subscribers. Hence, λ and $\bar{\theta}$ are identified by separate features of the data.

The supply-side price constraints provide identification of the market size, M. To see why, consider estimation of the demand model without price constraints. In this case we would simply assume a market size, estimate all other parameters using demand-side information only. While this approach can deliver a good fit of the observed market shares, there is no assurance that the optimal prices for the estimated demand model will be equal to the observed prices. In fact, if we set the market size to 100,000 and estimate using demand-side moments only, we compute predicted optimal prices that are significantly less than the observed prices. This suggests α is over-estimated—consumers are too sensitive to price.

By reducing the stipulated market size, the actual market share of inside goods increases. This implies the estimate for α must decrease in order to predict higher probabilities of purchase.[48] Importantly, optimal prices depend on α but not M. Hence, lowering M leads to a lower estimate of α, which in turn leads to higher predicted optimal prices. We can therefore estimate M by incorporating an optimal pricing constraint in the estimation. This is why the supply-side pricing constraints provide identification of the market size, and allow us to fit the demand moments while generating reasonable predicted prices.

A final comment on the flexibility of the model. Notice from the last two columns of Table 9 that the rank-ordering of play popularity is different for non-subscribers (mainly single-play buyers) than it is for subscribers (which is driven by the tastes of *pick-5* buyers because full season buyers attend all plays). Our specification can explain this difference in the following way. The variance terms, $\Sigma(k,k)$, explain the relative popularity of plays among the single-ticket buyers. The covariance terms in Σ explain the popularity of certain pairings by the *pick-5* buyers, which also helps to explain differences in the popularity of plays that are contrary to

[48]This reasoning suggests α and M are not separately identified from demand-side moments alone. From a practical standpoint this is right, since rich variation in the data is needed for separate identification. Formally, however, α and M are separately identified from demand moments, based on functional form.

the rankings of the single-ticket buyers. In other words, allowing for covariance in tastes gives us the flexibility to explain differences in play shares between single-play buyers and multi-play buyers.

4.4 Results

The parameter estimates for the structural demand model are presented in Table 12a. (Standard errors for the variance-covariance matrix are reported in Table 12b.) Estimates of the variance coefficients from the distribution of ϵ vary from 1.00 to 3.18. The estimates for the covariances of ϵ vary from 0.78 to 1.91. 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 than for the 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 .60 and .97 (the mean correlation coefficient is .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 estimate for consumers' sensitivity to price (α) is 4.60. To compute the implied price elasticity of aggregate demand, we increase the price of all possible bundles of tickets by 1% and measure the change in total tickets sold to all plays. The resulting price elasticity of aggregate demand is 1.11. Interestingly, when we implement this calculation the demand for certain mid-sized bundles actually increases, despite increasing price for those bundles. Intuitively, a 1% increase in all prices causes some consumers to substitute away from large to smaller-sized bundles. For example, demand for the Lucie Stern combination of 3 plays increases by 43% in response to a 1% increase in all prices.

The estimated probability of an individual being a theater-lover is .077, and the estimated market size is 37,435. We estimate that theater-lovers' utility for any single play is higher than for regular consumers by an amount equal to 2.05 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% of non-theater-lovers choose the outside option, while the predicted proportions choosing 1, 2, 3, or 4 individual goods are 12%, 4.3%, 1.7%, and 0.17%. Among this same group of consumers, the predicted market shares for the Lucie Stern, *pick-5*, and all-8 combinations are 0.75%, 8.6%, and 9.5%. For theater-lovers, on the other hand, only 8% choose the outside good, while 13% choose a bundle of 5 (accounting for 11% of *pick-5* purchases), and 68% subscribe to all 8 plays (accounting for 37% of all-8 purchases).

The non-monotonicity 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 either like 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.

To evaluate the fit of the estimated model, we compare various measures of actual and predicted market shares. We find a reasonably close fit between the actual and predicted overall market shares for different bundle sizes, where we compute the actual shares using the estimated market size as the denominator. The actual fraction that choose to be *pick-5* subscribers is 7.5%, and we predict 8.9%. The actual fraction all-8 subscribers is 13.7%, and we predict 14.0%.

In Table 13 we present actual and predicted shares of sales for each play. In the top portion of the table we show the play shares for all consumers (ignoring the outside option). Even though the predicted shares vary across plays from 19.5% to 29.3%, the actual and predicted shares are all within 1.5 percentage points, except for the first play. The fit is more uneven when we condition the play shares on how they are bundled. The predicted conditional shares nevertheless capture a key feature of the data: market shares are quite skewed for unbundled sales, while being much more symmetric for the *pick-5* subscribers.[49] For example, the most popular show among single-play purchases is play (4) with an observed share of 37.7%, more than four times that of play (2) which is the least popular show among single-play buyers. By contrast, among *pick-5* sales, the most popular show, play (8)—with a share of 14.6%—is purchased only slightly more often than the least popular show, play (3)—at 10.3%.

The estimated variances and covariances play a critical role here. Note, for example, that

[49]By construction, the play shares must be identical when conditioned on being bundled as part of the all-8 package.

the market share for play (4) among unbundled sales is high relative to its popularity among pick-5 sales. The model explains this fact in part by estimating a large variance for play (4) in conjunction with relatively low correlations with all the other plays: the high variance together with low correlations imply that play (8) will have high demand individually, but will tend not to be purchased together with other plays.

4.5 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 examine how particular changes in the model affect the relative profits of these different pricing structures.

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 3-play bundle, a discount for choosing any 5 plays (*pick-5*), and a discount for the bundle of all 8 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 8 plays. In estimation these two predicted optimal prices exactly match the observed prices (to within 1e-4). 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.

Table 14 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%). In this case the variability in profits across price structures is somewhat low compared to many of the simulations in Section 3. 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 8 goods, PB performs quite badly. Even UP is more profitable than PB in this setting (by 5%). This reinforces the point that PB is not necessarily a good option for firms.

Focusing on BSP in Table 14, we find that: (i) BSP attains 0.9% higher profit than CP, and (ii) BSP attains 98.5% 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* 8 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 .60 and .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 8 plays)—a discount of 42% 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 there are 9.9% of consumers that attend exactly one play, while under BSP only 2.6% of consumers attend just one play. Under CP there are 9.0% of consumers that attend all 8 plays, while under BSP 12.6% 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 four—under UP play four has the highest level of sales, and under CP play four has the highest optimal price. For CP we compute that 25.2% of consumers attend play four, and by construction every ticket is sold at the price \$44.08. Under BSP we find that 27.1% consumers attend play four. The average (per play) price paid by consumers that attend play four under BSP is \$36.70. It follows that BSP obtains 21% less revenue from play four than does CP.[50] 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% more revenue for play one (the lowest-demand

[50] 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 of profits.

good) than CP.

The fact that CP generates more revenue than BSP for the popular good is not a general result. For other parameterizations of demand we may find the reverse. It is interesting to note that under BSP there is price dispersion among the consumers that attend play four (or any other particular play), and the amount of surplus that the firm extracts from consumers of popular plays depends on whether these consumers pay higher or lower prices, on average. In our simulation, under BSP 46% of consumer that attend play four did so as part of the full-season subscription, for whom the per-play ticket price is \$32.89. At the other extreme, for 8% of the attendees of play four this was the only play they attended, and hence the ticket price for them is \$56.41. And there are six other price points in between, depending upon how many other plays the consumer purchased tickets for. The distribution of ticket prices paid by attendees is different for each play under BSP. For example, under BSP nobody buys a ticket to play one by itself, and hence the highest price paid by any individual attending play one is \$46.92 (i.e. per-play price for a two-good bundle).

The within-play price variation that arises under BSP is a consequence of people self-selecting different size bundles. Sorting occurs along the dimension of the total number of plays. That is, consumers with a high aggregate valuation for multiple plays tend to purchase larger-sized bundles and pay a higher total price. But this sorting under BSP does not necessarily imply that in a given play, consumers with a high valuation for that play will pay a high price for attending that play. Consider the consumers that attend play four under BSP: the correlation between consumers' valuations of play four and each consumers' per-play price is -0.12. Intuitively, this is due to the high degree of positive correlation in the estimated demand system—consumers that like play four also tend to like many other plays, and therefore tend to obtain discounts for buying larger bundle sizes. For the same reasoning, if demand system exhibited a high degree of negative correlation, we would expect the within-play correlation between price and valuations to be positive. It follows that the less positive correlation there is in the demand system, the more surplus BSP extracts from higher-demand goods, relative to CP.

As expected, we predict optimal prices under TW that are very close the actual prices set by *TheatreWorks*.[51] 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. Our estimated model indicates more aggressive discounting than *TheatreWorks*' actual price schedule, but not by much. The TW price structure appears

[51] To understand why optimal prices are not exactly equal to actual prices, recall that in the estimation, we only impose the optimality of the individual-play and all-8 prices, rather than the full set of prices.

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% 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 spreads of the estimated variances (holding all other parameters fixed), and recomputed the optimal prices and profits under the various pricing strategies.[52] We find that BSP remains more profitable than CP even when the highest price for a play under CP ($95) is 340% greater than the lowest price play ($28). However, if the price difference increases to over 385% (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

[52] 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.

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.

Capacity constraints, which we have so far assumed away, could also favor CP over BSP. To assess the impact of capacity constraints, we assume capacity is equal for all plays, and is set to $\Delta\%$ of the predicted sales for the most popular play under UP (where UP is computed under the assumption there are no capacity constraints). How small must Δ be (i.e., how tight does the capacity constraint have to be) for CP to be more profitable than BSP? We find that if $\Delta = 90$, BSP is still more profitable than CP, but if we lower it to 80 then CP becomes more profitable than BSP. Hence, it appears the greater is excess demand (i.e. the lower is capacity relative to demand) the more likely that CP is more profitable than BSP. Regardless, for *TheatreWorks* (as well as many other firms) the capacity for each good is to some extent endogenous. This makes it less likely that capacity constraints are a reason to prefer CP over BSP when capacity is a choice variable for firms.[53]

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 re-compute 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% more profitable than CP.

Finally, we argued in Section 3 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:

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

where n_j equals the number of goods in bundle j and γ is a parameter. We set $\gamma = -.2$ to capture diminishing marginal utility, and compute optimal prices holding all other parameters fixed at the estimated values under the baseline model.[54] Unsurprisingly the profits under all pricing schemes are lowered relative to the baseline. But now BSP attains 8.8% higher profit than CP, compared to 0.9% in the baseline model. Hence, the inclusion of diminishing marginal utility can *increase* the profits of BSP relative to CP.

As we found in the numerical experiments discussed in Section 3, the above perturbations to

[53] In fact *TheatreWorks* offers more performances for some shows than for others, suggesting they choose capacities for each play (see Table 9).

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

the estimated demand model demonstrate that in some cases BSP is more profitable than CP, and in other cases the reverse is true. What is striking about the perturbations we analyze here is that it requires rather dramatic changes to the estimated model for CP to out-perform BSP. That is to say, the relative profitability of BSP seems very robust in our analysis.

5 Conclusion

We have examined the profitability of several incomplex pricing strategies for multiproduct firms, 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, bundle-size pricing tends to attain nearly the same level of profits as mixed bundling in a broad range of demand and cost scenarios. Hence, mixed bundling involves considerable redundancy—it includes many prices that are of negligible importance to profitability. Second, bundle-size pricing tends to be more profitable than component pricing, even in circumstances with a high degree of demand asymmetry across products.

To illustrate the empirical relevance of our findings we estimate the demand facing a theater company that produces a season of 8 plays, and compute the profitability of each pricing scheme in this case. We find that bundle-size pricing is 0.9% more profitable than component pricing, and bundle-size pricing attains 98.5% 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.

Our results represent a significant push towards understanding the merits of feasible pricing schemes for multiproduct firms. What insight does the prior literature on bundling have for a firm with 5 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 suggests specific advice to such a firm; bundle-size pricing (5 prices) tends to attain about 99% of the mixed bundling profit, and is almost certainly more profitable than either component pricing or pure bundling.

An important theme of our findings is that bundling-like 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 9 games) more often than they employ component pricing strategies (such as charging prices that vary by opponent or by day of the week).[55] Also, online music sellers almost never charge different prices for different music tracks, even though demand is dramatically stronger for some songs than others. But music is sold via subscriptions (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.[56]

More generally, our results suggest an additional explanation for the observed simplicity of multiproduct firms' pricing strategies. Other authors have proposed various theories to explain why, in practice, complex pricing schemes are costly to implement.[57] Our study is the first (to our knowledge) to quantify the benefits of complexity. Our findings suggest these benefits are generally small, so that the costs of complexity need not be large to make simplicity the best policy.

[55]We examined the pricing for all 30 major league teams during the 2006 season. 16 teams employed some form of bundling (not including season-ticket subscriptions), whereas only 7 charged prices that varied by opponent or by day of the week.

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

[57]See, for example, Blinder, Canetti, Lebow and Rudd (1998) and Kahneman, Knetsch and Thaler (1986).

References

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

Allenby, G.M. and P.E. Rossi (1999): "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89, 57–78.

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

Bajari, P.L. and C.L. Benkard (2005): "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, 113(6), 1239–76.

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

Berry, S., J. Levinsohn and A. Pakes (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, 112(1), 68–104.

Blinder, A.S., E.R.D. Canetti, D.E. Lebow and J.B. Rudd (1998): "Asking About Prices: A New Approach to Understanding Price Stickiness," *Working Paper, Russell Sage Foundation.*

Crawford, G.S. (2006): "The Discriminatory Incentives to Bundle in the Cable Television Industry," forthcoming *Quantitative Marketing and Economics.*

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

Fang, H. and P. Norman (2006): "To Bundle or Not to Bundle," forthcoming in *RAND Journal of Economics.*

Geng, X., M.B. Stinchcombe, and A.B. Whinston (2005): "Bundling Information Goods of Decreasing Value," *Management Science*, 51(4), 662–7.

Gill, P., W. Murray and M. Saunders (2002): "SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization," *Society for Industrial and Applied Mathematics Journal of Optimization*, 12(4), pp. 979–1006.

Hajivassiliou, V., D. McFadden, and P. Ruud (1996): "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational Results," *Journal of Econometrics*, 72, 85–134.

Hartmann, W.R., and V.B. Viard (2007): “Do Frequency Reward Programs Create Switching Costs? A Dynamic Structural Analysis of Demand in a Reward Program,” forthcoming *Quantitative Marketing and Economics*.

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

Jedidi, K., S. Jagpal and P. Manchanda (2003): “Measuring Heterogeneous Reservation Prices for Product Bundles,” *Marketing Science*, 22(1), 107–30.

Kahneman, D., J.L. Knetsch and R. Thaler (1986): “Fairness as a Constraint on Profit Seeking: Entitlements in the Market,” *American Economic Review*, 76(4), 728–41.

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

Leslie, P., and A. Sorensen (2007): “Ticket Resale,” *Mimeo*.

Malueg, D. and C. Snyder (2006): “Bounding the Relative Profitability of Price Discrimination,” *International Journal of Industrial Organization*, 24(2), 995–1011.

Manelli, A.M. and D.R. Vincent (2006): “Bundling as an Optimal Selling Mechanism for a Multi-Good Monopolist,” *Journal of Economic Theory*, 127, 1–35.

McAfee, R.P. (2002): “Coarse Matching,” *Econometrica*, 70(5), 2025–34.

McAfee, R.P., J. McMillan and M. Whinston (1989): “Multiproduct Monopoly, Commodity Bundling, and Correlation of Values,” *Quarterly Journal of Economics*, 114, 371–84.

McFadden, D. (1989): “A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration,” *Econometrica*, 57, 995–1026.

Pakes, A., and D. Pollard (1989): “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 57, 1027–1057.

Rogerson, W.P. (2003): “Simple Menus of Contracts in Cost-Based Porcurement and Regulation,” *American Economic Review*, 93(3), 919–26.

Schmalensee, R. (1984): “Gaussian Demand and Commodity Bundling,” *Journal of Business*, 57(1), 211–30.

Stigler, G.J. (1963): “United States v. Loew’s Inc.: A Note on Block Booking,” *Supreme Court Review*, 152–7.

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

Wilson, R.B. (1993): *Nonlinear Pricing*, Oxford University Press.

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 m^{th}-highest valuation (i.e., the m^{th} 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

$$\textit{(i)} \qquad 0 < \Delta < c$$

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

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

$$\begin{aligned} \tilde{\pi}(\Delta) \quad &= \quad (\,p^* - Kc\,) \int_{\Delta}^{\infty} \int_{p^*-r}^{\infty} g(y,r)dydr \quad + \\ &\qquad (\,p^* - \Delta - (K-1)c\,) \int_0^{\Delta} \int_{p^*-\Delta}^{\infty} g(y,r)dydr \end{aligned}$$

The difference from the PB profits is then

$$\tilde{\pi}(\Delta) - \tilde{\pi}(0) \quad = \quad (c - \Delta) \int_0^{\Delta} \int_{p^*-r}^{\infty} g(y,r)dydr + (\, p^* - \Delta - (K-1)c\,) \int_0^{\Delta} \int_{p^*-\Delta}^{p^*-r} g(y,r)dydr$$

Conditions (i) and (ii) of the proposition simply guarantee that this difference is 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 3.2. 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 $\theta \leq 1.756739614$: $p_1^* = \frac{(1+\theta)}{3}$ $p_2^* = \frac{1}{3}\left(2 + 2\theta - \sqrt{2\theta^2 - 2\theta + 2}\right)$ Otherwise: $p_1^* = p_2^* = \frac{1}{4} + \frac{\theta}{2}$	If $\theta \leq 1.756739614$: $\pi^* = \frac{\left(2\theta^2 - 2\theta + 2\right)^{3/2} - 3\theta^3 + 9\theta^2 + 9\theta - 3}{27\theta}$ 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 $\theta > 2$: $\pi^* = \frac{27\theta^2 + 36\theta - 4}{108\theta}$

Appendix C

Available on request from the authors (or via our home pages on the web). It's very long.

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 different prices
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

Table 2. Alternative taste distributions

Name	Description
Exponential	v_{ik}'s are independent exponential random variables with means between 0.2 and 2.0
Logit	v_{ik}'s are independent extreme value random variables with means between 0 and 2.5, and scale parameter = 0.25
Lognormal	v_{ik}'s are independent lognormal random variables; $log(v_{ik})$ has variance 0.25 and mean between -1.5 and 1
Lognormal(-)	$log(v_i)$ is a multivariate normal random vector with means ranging between -1.5 and 1, and negative pairwise correlations between products*
Lognormal(+)	$log(v_i)$ is a multivariate normal random vector with means ranging between -1.5 and 1, and positive pairwise correlations between products
Normal	v_{ik}'s are independent normal random variables with variances equal to 0.25, and means between -1 and 2.5
Normal(-)	v_i is a multivariate normal random vector with means ranging between -1 and 2.5, and negative pairwise correlations between products*
Normal(+)	v_i is a multivariate normal random vector with means ranging between -1 and 2.5, and positive pairwise correlations between products
Normal(+/-)	v_i is a multivariate normal random vector with means ranging between -1 and 2.5, and pairwise correlations are a mix of positive and negative values**
Normal(v)	v_{ik}'s are independent normal random variables with means equal to zero, and variances between 0.25 and 1.75
Normal(v-)	v_i is a multivariate normal random vector with means equal to zero, variances between 0.25 and 1.75, and negative pairwise correlations between products*
Normal(v+)	v_i is a multivariate normal random vector with means equal to zero, variances between 0.25 and 1.75, and positive pairwise correlations between products
Uniform	v_{ik}'s are independent uniform random variables on $[0, a_k]$, with a_k between 0.4 and 4

* For distributions with negatively correlated tastes, we set the pairwise correlation coefficients all equal to $\underline{r}/2$, where $\underline{r}$ is the smallest (i.e., most negative) value such that the covariance matrix remains positive definite. For K=(2, 3, 4, 5) the correlation coefficients are (-0.5, -0.25, -0.1667, -0.125) respectively.

** For $K > 2$, we assume tastes for one pair of products have a correlation of -.25 and for another pair of products 0.25.

Table 3. Percentiles of profits (as a fraction of BSP profits) with zero marginal costs

		Pricing Scheme			
Taste Distn.	percentile	UP	CP	PB	MB
	.01	0.632	0.724	0.995	1.000
Exponential	.50	0.752	0.812	1.000	1.001
	.99	0.878	0.977	1.000	1.041
	.01	0.532	0.838	0.956	1.000
Logit	.50	0.779	0.899	0.989	1.003
	.99	0.956	0.984	1.000	1.019
	.01	0.403	0.734	0.999	1.000
Lognormal	.50	0.698	0.809	1.000	1.000
	.99	0.861	0.945	1.000	1.002
	.01	0.623	0.806	0.972	1.000
Normal	.50	0.831	0.900	0.995	1.001
	.99	0.985	1.023	1.000	1.043
	.01	0.690	0.930	0.901	1.000
Normal(+)	.50	0.900	0.962	0.978	1.000
	.99	0.992	1.024	1.000	1.081
	.01	0.454	0.544	0.937	1.000
Normal(-)	.50	0.671	0.743	1.000	1.000
	.99	0.991	1.025	1.000	1.090
	.01	0.837	0.872	0.951	1.000
Normal(v)	.50	0.895	0.936	0.973	1.022
	.99	0.961	1.039	0.997	1.097
	.01	0.419	0.800	0.982	1.000
Uniform	.50	0.777	0.884	0.998	1.022
	.99	0.919	1.003	1.000	1.072

The table reports percentiles of the ratio of profits to BSP profits, calculated across roughly 900 experiments represented in each cell. For a detailed summary of the price strategies and the taste distributions see Tables 1 and 2, respectively.

Table 4. Percentiles of profits (as a fraction of BSP profits) with equal positive marginal costs

		Pricing Scheme			
Taste Distn.	percentile	UP	CP	PB	MB
	.01	0.720	0.758	0.788	1.000
Exponential	.50	0.812	0.859	0.975	1.004
	.99	0.946	1.001	0.994	1.048
	.01	0.552	0.840	0.434	1.000
Logit	.50	0.801	0.906	0.958	1.003
	.99	0.991	0.998	0.995	1.019
	.01	0.563	0.728	0.841	1.000
Lognormal	.50	0.743	0.828	0.995	1.000
	.99	0.935	0.971	1.000	1.004
	.01	0.632	0.809	0.344	1.000
Normal	.50	0.851	0.907	0.929	1.001
	.99	1.000	1.001	1.000	1.011
	.01	0.700	0.932	0.417	1.000
Normal(+)	.50	0.915	0.966	0.898	1.000
	.99	0.999	1.000	0.999	1.009
	.01	0.461	0.539	0.323	1.000
Normal(-)	.50	0.700	0.768	0.969	1.000
	.99	1.000	1.020	1.000	1.021
	.01	0.864	0.892	0.676	1.000
Normal(v)	.50	0.912	0.956	0.804	1.022
	.99	0.970	1.056	0.902	1.101
	.01	0.592	0.820	0.746	1.000
Uniform	.50	0.823	0.924	0.969	1.031
	.99	0.960	1.032	0.993	1.086

The table reports percentiles of the ratio of profits to BSP profits, calculated across the roughly 900 experiments represented in each cell. For a detailed summary of the price strategies and the taste distributions see Tables 1 and 2, respectively. Marginal cost is set to 0.2 for all products.

Table 5. Percentiles of profits (as a fraction of BSP profits) with unequal marginal costs

Taste Distn.	percentile	Pricing Scheme UP	CP	PB	MB
	.01	0.657	0.824	0.812	1.000
Exponential	.50	0.839	0.940	0.910	1.054
	.99	0.939	1.058	0.988	1.172
	.01	0.373	0.834	0.829	1.000
Logit	.50	0.713	0.903	0.950	1.014
	.99	0.949	0.992	0.989	1.070
	.01	0.438	0.774	0.957	1.000
Lognormal	.50	0.720	0.888	0.989	1.034
	.99	0.911	0.993	0.997	1.081
	.01	0.466	0.799	0.798	1.000
Normal	.50	0.792	0.910	0.951	1.009
	.99	0.981	1.097	1.000	1.143
	.01	0.000	0.931	0.809	1.000
Normal(+)	.50	0.841	0.967	0.932	1.003
	.99	0.988	1.035	1.000	1.160
	.01	0.303	0.523	0.838	1.000
Normal(-)	.50	0.632	0.768	0.987	1.007
	.99	1.000	1.122	1.000	1.146
	.01	0.834	0.894	0.760	1.000
Normal(v)	.50	0.908	0.967	0.823	1.032
	.99	0.969	1.096	0.933	1.149
	.01	0.437	0.864	0.795	1.000
Uniform	.50	0.813	0.999	0.893	1.101
	.99	0.947	1.091	0.992	1.230

The table reports percentiles of the ratio of profits to BSP profits, calculated across the roughly 900 experiments represented in each cell. For a detailed summary of the price strategies and the taste distributions see Tables 1 and 2, respectively. Marginal cost equals 0.5 times consumers' mean product valuation.

Table 6. Percentiles of profits (as a fraction of BSP profits) with capacity constraints

		Pricing Scheme			
Taste Distn.	percentile	UP	CP	PB	MB
Exponential	.01	0.455	0.818	0.652	1.000
	.50	0.859	0.947	0.857	1.069
	.99	1.227	1.246	1.000	1.303
Logit	.01	0.436	0.858	0.743	1.000
	.50	0.770	0.922	0.986	1.004
	.99	1.000	1.025	1.000	1.071
Lognormal	.01	0.351	0.746	0.868	1.000
	.50	0.700	0.831	1.000	1.000
	.99	0.934	0.953	1.000	1.061
Normal	.01	0.583	0.821	0.666	1.000
	.50	0.833	0.915	0.987	1.000
	.99	1.004	1.043	1.000	1.085
Normal(+)	.01	0.597	0.907	0.687	1.000
	.50	0.876	0.964	0.970	1.000
	.99	1.005	1.038	1.000	1.056
Normal(-)	.01	0.482	0.667	0.704	1.000
	.50	0.774	0.852	0.990	1.010
	.99	1.001	1.054	1.000	1.128
Normal(v)	.01	0.847	0.919	0.499	1.000
	.50	0.912	0.976	0.709	1.027
	.99	1.125	1.125	0.949	1.138
Uniform	.01	0.336	0.858	0.729	1.000
	.50	0.844	0.977	0.914	1.064
	.99	1.143	1.198	1.000	1.285

The table reports percentiles of the ratio of profits to BSP profits, calculated across the roughly 900 experiments represented in each cell. Marginal costs are zero, but there is a binding capacity constraint which varies by experiment. It is set by first computing the optimal uniform price under no capacity constraints, finding the quantity demanded for the most popular product, and setting the constraint equal to 0.9 times that quantity.

Table 7. Price differences and market shares, by bundle size

K=3

Bundle size	Average price differences		Market shares		
	$\lvert p_{CP} - p_{MB}\rvert / p_{MB}$	$\lvert p_{BSP} - p_{MB}\rvert / p_{MB}$	CP	BSP	MB
1	0.295	0.657	0.384	0.125	0.140
2	0.146	0.183	0.227	0.170	0.197
3	0.174	0.037	0.082	0.287	0.274

K=4

Bundle size	Average price differences		Market shares		
	$\lvert p_{CP} - p_{MB}\rvert / p_{MB}$	$\lvert p_{BSP} - p_{MB}\rvert / p_{MB}$	CP	BSP	MB
1	0.359	0.809	0.327	0.089	0.097
2	0.196	0.306	0.250	0.134	0.157
3	0.167	0.123	0.140	0.147	0.169
4	0.202	0.037	0.051	0.257	0.238

K=5

Bundle size	Average price differences		Market shares		
	$\lvert p_{CP} - p_{MB}\rvert / p_{MB}$	$\lvert p_{BSP} - p_{MB}\rvert / p_{MB}$	CP	BSP	MB
1	0.409	0.929	0.275	0.062	0.067
2	0.238	0.403	0.245	0.106	0.119
3	0.185	0.197	0.170	0.119	0.144
4	0.183	0.094	0.092	0.139	0.154
5	0.218	0.035	0.035	0.232	0.212

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 14.0% of consumers to purchase a single product.

Table 8. Average welfare effects

		CP	BSP	MB
$K = 3$	Total output	1.086	1.329	1.359
	Consumer surplus	0.590	0.526	0.523
	Producer surplus	0.929	1.072	1.089
	Total surplus	1.520	1.597	1.611
	Dead weight loss	0.367	0.290	0.276
$K = 4$	Total output	1.452	1.827	1.874
	Consumer surplus	0.796	0.692	0.670
	Producer surplus	1.261	1.489	1.516
	Total surplus	2.057	2.180	2.186
	Dead weight loss	0.493	0.370	0.364
$K = 5$	Total output	1.815	2.347	2.410
	Consumer surplus	1.000	0.850	0.808
	Producer surplus	1.590	1.913	1.951
	Total surplus	2.589	2.763	2.760
	Dead weight loss	0.619	0.445	0.448

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

Table 9. Summary of ticket sales

Play	Type	Number of Performances	Average Attendance	Ticket sales (subscription)	Ticket sales (non-subscription)
A Little Night Music	Musical	30	294.87	7018	1828
All My Sons	Drama	33	233.85	6826	891
Bat Boy	Musical	30	263.93	6782	1136
Memphis	Musical	30	352.40	6999	3573
My Antonia	Drama	26	312.38	7002	1120
Nickel and Dimed	Drama	26	343.62	6800	2134
Proof	Drama	25	319.88	6885	1112
The Fourth Wall	Comedy	29	313.83	7385	1716
Total		229	302.21	55,697	13,510

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 5 were performed at the Mountain View Center for the Performing Arts (capacity=589).

Table 10. Sales by purchase option

Purchase option	Price per play ($)	Number of consumers
Non-subscription:		
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

For non-subscription 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 non-subscription 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.

Table 11. Correlation of tastes for *pick-5* subscribers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) *A Little Night Music*	.000							
(2) *All My Sons*	-.026	.000						
(3) *Bat Boy*	.067	-.233	.000					
(4) *Memphis*	.072	-.081	.257	.000				
(5) *My Antonia*	.177	.067	-.086	-.037	.000			
(6) *Nickel and Dimed*	-.160	-.009	-.013	-.039	.001	.000		
(7) *Proof*	-.066	.210	-.030	-.057	-.094	.008	.000	
(8) *The Fourth Wall*	-.038	.034	.003	-.117	-.007	.196	.008	.000

This is the difference between the observed correlation matrix and the correlation matrix that would be expected if plays were chosen independently (i.e., no correlation in tastes). It is constructed for the 8,005 purchasers of the flexible 5-play subscription.

Table 12a. Estimated coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Covariances (Σ)				
(1)	1.0000							
(2)	0.9340	1.2170						
(3)	1.2114	1.3722	1.6579					
(4)	0.9821	1.3339	1.4611	3.1798				
(5)	0.8743	1.1015	1.1842	1.4908	1.3487			
(6)	1.1672	1.3429	1.5932	1.9128	1.3214	1.9680		
(7)	0.7808	1.0502	1.1710	1.7798	1.1236	1.4368	1.4775	
(8)	1.1159	1.2817	1.5560	1.7022	1.2872	1.7066	1.6636	2.2433
	Estimate	Std. error						
Price sensitivity (α)	4.6009	(0.0598)						
Probability of theater-lover (λ)	0.0771	(0.0133)						
Increment for theater-lovers ($\bar{\theta}$)	2.0519	(0.2108)						
Market size	37435	(809)						

Standard errors for Σ are in Table 12a. All parameter estimates are significant at the 1% level.

Table 12b. Standard errors for estimated covariances (Σ) in Table 12a

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(2)	0.0397	0.0136						
(3)	0.0191	0.0395	0.0193					
(4)	0.0450	0.0448	0.1028	0.0540				
(5)	0.0313	0.0285	0.0385	0.1009	0.0178			
(6)	0.0237	0.0318	0.0406	0.0501	0.0514	0.0264		
(7)	0.0321	0.0291	0.0349	0.0466	0.0335	0.0629	0.0203	
(8)	0.0373	0.0337	0.0452	0.0526	0.0415	0.0410	0.0619	0.0338

Table 13. Actual and predicted market shares of each play

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Shares among all consumers (%)				
Actual	23.6	20.6	21.1	28.2	21.7	23.9	21.4	24.3
Predicted	19.5	20.7	22.5	29.3	21.4	23.9	21.6	25.4
				Shares among all unbundled sales (%)				
Actual	19.3	9.40	12.0	37.7	11.8	22.5	11.7	18.1
Predicted	5.32	2.78	6.29	37.6	8.83	11.1	10.0	18.1
				Shares among *pick-5* subscribers (%)				
Actual	13.5	10.6	10.3	13.3	13.3	11.9	12.5	14.6
Predicted	9.25	11.8	13.9	12.8	11.4	15.8	11.3	13.7

Table 14. 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

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 8 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 3 plays, p_5 is the per-play price for any combination of 5 plays, and p_8 is the per-play price if you buy all 8. For CP, p_1-p_8 are the prices for the 8 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 2-play bundles, and so forth). The revenue and consumer surplus numbers are normalized by the market size—i.e., we report revenue per consumer.

Figure 1: Separation of consumers under CP and BSP

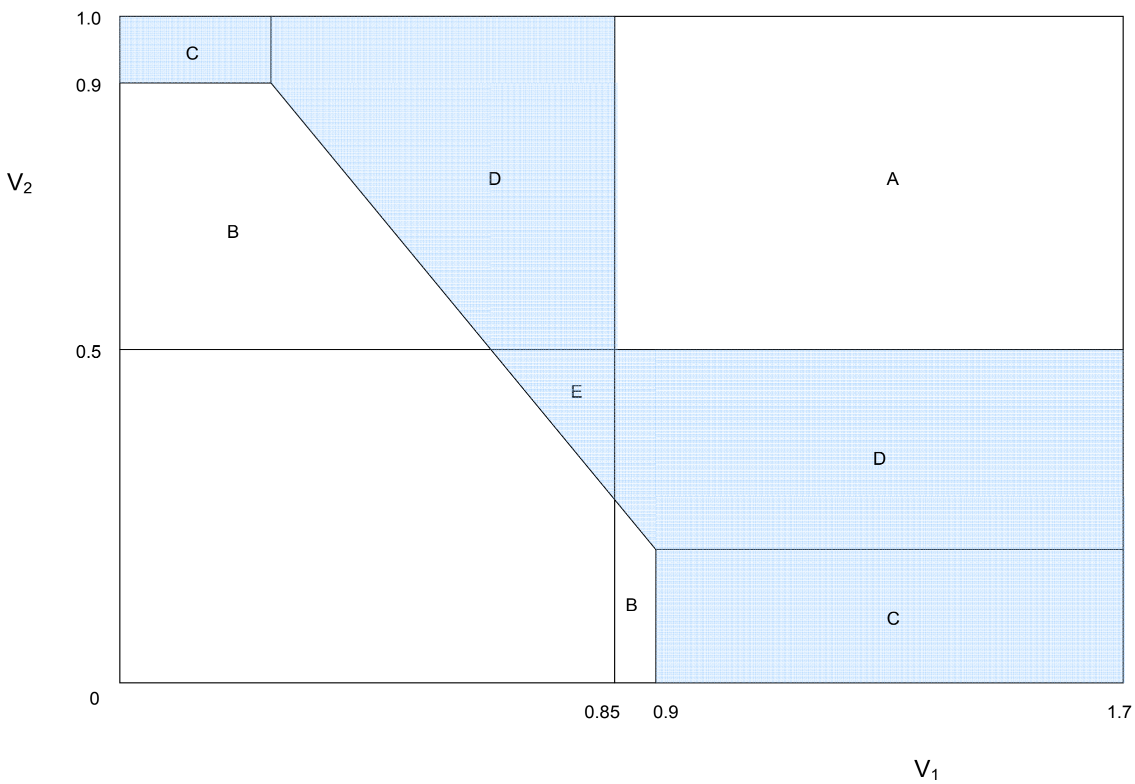
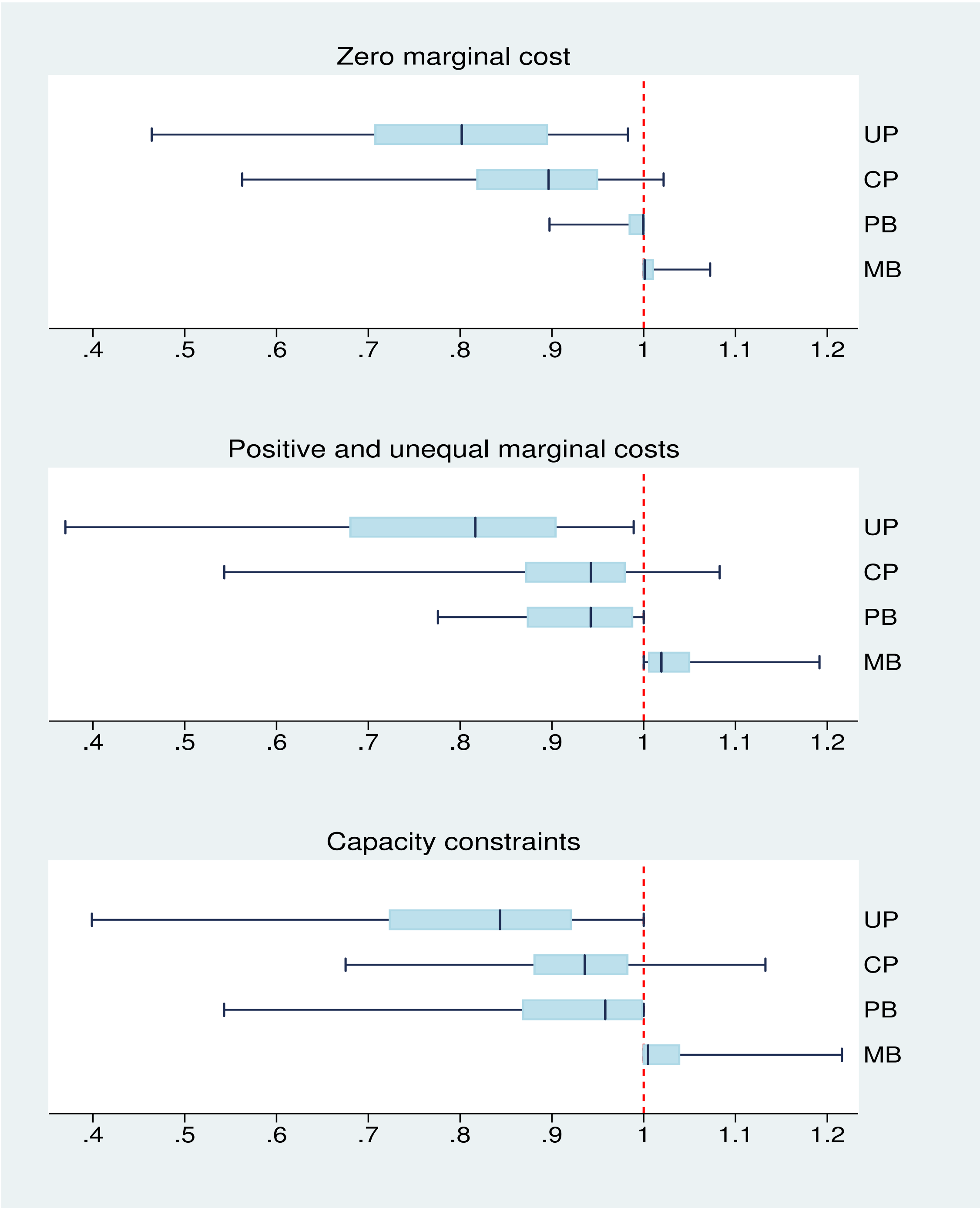


Figure 2: Distributions of profits for each pricing strategy, relative to BSP, under different assumptions on marginal costs



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