

PRICE DISCRIMINATION AND COMPETITION

LARS A. STOLE

University of Chicago, GSB

Contents

Abstract	2223
Keywords	2223
1. Introduction	2224
2. First-degree price discrimination	2229
3. Third-degree price discrimination	2231
3.1. Welfare analysis	2231
3.2. Cournot models of third-degree price discrimination	2233
3.3. A tale of two elasticities: best-response symmetry in price games	2234
3.4. When one firm's strength is a rival's weakness: best-response asymmetry in price games	2239
3.5. Price discrimination and entry	2244
3.6. Collective agreements to limit price discrimination	2246
4. Price discrimination by purchase history	2249
4.1. Exogenous switching costs and homogeneous goods	2251
4.2. Discrimination based on revealed first-period preferences	2254
4.3. Purchase-history pricing with long-term commitment	2257
5. Intrapersonal price discrimination	2259
6. Non-linear pricing (second-degree price discrimination)	2262
6.1. Benchmark: monopoly second-degree price discrimination	2264
6.2. Non-linear pricing with one-stop shopping	2267
6.2.1. One-dimensional models of heterogeneity	2267
6.2.2. Multidimensional models of heterogeneity	2271
6.3. Applications: add-on pricing and the nature of price-cost margins	2275
6.4. Non-linear pricing with consumers in common	2277
6.4.1. One-dimensional models	2278
6.4.2. Multidimensional models	2280
7. Bundling	2281
7.1. Multiproduct duopoly with complementary components	2282
7.2. Multiproduct monopoly facing single-product entry	2284
8. Demand uncertainty and price rigidities	2286

8.1. Monopoly pricing with demand uncertainty and price rigidities	2288
8.2. Competition with demand uncertainty and price rigidities	2290
9. Summary	2292
Acknowledgements	2292
References	2292

Abstract

This chapter surveys the developments in price discrimination theory as it applies to imperfectly competitive markets. Broad themes and conclusions are discussed in the areas of first-, second- and third-degree price discrimination, pricing under demand uncertainty, bundling and behavior-based discrimination.

Keywords

Price discrimination, Oligopoly, Imperfect competition, Market segmentation, Demand uncertainty, Bundling

JEL classification: D400, L100, L400, L500

1. Introduction

Firms often find it profitable to segment customers according to their demand sensitivity and to price discriminate accordingly. In some settings, consumer heterogeneity can be directly observed, and a firm can base its pricing upon contractible consumer characteristics. In other settings, heterogeneity is not directly observable but can be indirectly elicited by offering menus of products and prices and allowing consumers to self-select. In both cases, the firm seeks to price its wares as a function of each consumer's underlying demand elasticity, extracting more surplus and increasing sales to more elastic customers in the process.

When the firm is a monopolist with market power, the underlying theory of price discrimination is now well understood, as explained, for example, by Varian (1989) in an earlier volume in this series.¹ On the other extreme, when markets are perfectly competitive and firms have neither short-run nor long-run market power, the law of one price prevails and price discrimination cannot exist.² Economic reality, of course, largely lies somewhere in between the textbook extremes, and most economists agree that price discrimination arises in oligopoly settings.³ This chapter explores price discrimination in these imperfectly competitive markets, surveying the theoretical literature.⁴

Price discrimination exists when prices vary across customer segments in a manner that cannot be entirely explained by variations in marginal cost. Stigler's (1987) definition makes this precise: a firm price discriminates when the ratio of prices is different

¹ In addition to Varian (1989), this survey benefited greatly from several other excellent surveys of price discrimination, including Philips (1983), Tirole (1988, ch. 3), Wilson (1993) and Armstrong (2006).

² It is straightforward to construct models of price discrimination in competitive markets without entry barriers in which firms lack long-run market power (and earn zero long-run economic profits), providing that there is some source of short-run market power that allows prices to remain above marginal cost, such as a fixed cost of production. For example, a simple free-entry Cournot model as discussed in Section 3.2 with fixed costs of production will exhibit zero long-run profits, prices above marginal cost, and equilibrium price discrimination. The fact that price discrimination can arise in markets with zero long-run economic profits suggests that the presence of price discrimination is a misleading proxy for long-run market power. This possibility is the subject of a recent symposium published in the *Antitrust Law Journal* (2003, vol. 70, No. 3); see the papers by Baker (2003), Baumol and Swanson (2003), Hurdle and McFarland (2003), Klein and Wiley (2003a, 2003b), and Ward (2003) for the full debate. Cooper et al. (2005) directly confront the issue of using price discrimination as proxy for market power by using a model in the spirit of Thisse and Vives (1988), which is discussed below in Section 3.4.

³ Much empirical work tests for the presence of price discrimination in imperfectly competitive environments. An incomplete sample of papers and markets includes: Shepard (1991) – gasoline service stations, Graddy (1995) – fish market, Goldberg (1995), Verboven (1996), Goldberg and Verboven (2005) – European automobiles, Leslie (2004) – Broadway theater, Busse and Rysman (2005) – yellow pages advertising, Clerides (2004), Cabolis et al. (2005) – books, Cohen (2001) – paper towels, Crawford and Shum (2001) – cable television, McManus (2004) – specialty coffee, Nevo and Wolfram (2002) – breakfast cereal, Besanko, Dube and Gupta (2003) – ketchup, Miravete and Röller (2003) – cellular communications. For the minority viewpoint that price discrimination is less common in markets than typically thought, see Lott and Roberts (1991).

⁴ Armstrong (2006) presents an excellent survey on these issues as well.

from the ratio of marginal costs for two goods offered by a firm.⁵ Such a definition, of course, requires that one is careful in calculating marginal costs to include all relevant shadow costs. This is particularly true where costly capacity and aggregate demand uncertainty play critical roles, as discussed in Section 8. Similarly, where discrimination occurs over the provision of quality, as reviewed in Section 6, operationalizing this definition requires using the marginal prices of qualities and the associated marginal costs.

Even with this moderately narrow definition of price discrimination, there remains a considerable variety of theoretical models that address issues of price discrimination and imperfect competition. These include classic third-degree price discrimination (Section 3), purchase-history price discrimination (Section 4), intrapersonal price discrimination (Section 5), second-degree price discrimination and non-linear pricing (Section 6), product bundling (Section 7), and demand uncertainty and price rigidities (Section 8). Unfortunately, we must prune a few additional areas of inquiry, leaving some models of imperfect competition and price discrimination unexamined in this chapter. Among the more notable omissions are price discrimination in vertical structures,⁶ imperfect information and costly search,⁷ the commitment effect of price discrimina-

⁵ Clerides (2002) contrasts Stigler's ratio definition with a price-levels definition (which focuses on the difference between price and marginal cost), and discusses the relevance of this distinction to a host of empirical studies.

⁶ For example, in a vertical market where a single manufacturer can price discriminate across downstream retailers, there are a host of issues regarding resale-price maintenance, vertical foreclosure, etc. As a simple example, a rule requiring that the wholesaler offer a single price to all retailers may help the upstream firm to commit not to flood the retail market, thereby raising profits and retail prices. While these issues of vertical markets are significant, we leave them largely unexplored in the present paper. Some of these issues are considered elsewhere in this volume; see, for example, the chapter by Rey and Tirole (2007).

⁷ The role of imperfect information and costly search in imperfectly competitive environments has previously received attention in the first volume of this series [Varian (1989, ch. 10, Section 3.4) and Stiglitz (1989, ch. 13)]. To summarize, when customers differ according to their information about market prices (or their costs of acquiring information), a monopolist may be able to segment the market by offering a distribution of prices, as in the model by Salop (1977), where each price observation requires a costly search by the consumer. A related set of competitive models with equilibrium price distributions has been explored by numerous authors including Varian (1980), Salop and Stiglitz (1977, 1982), Rosenthal (1980) and Stahl (1989). Unlike Salop (1977), in these papers each firm offers a single price in equilibrium. In some variations, firms choose from a continuous equilibrium distribution of prices; in others, there are only two prices in equilibrium, one for the informed and one for the uninformed. In most of these papers, average prices increase with the proportion of uninformed consumers and, more subtly, as the number of firms increases, the average price level can increase toward the monopoly level. These points are integrated in the model of Stahl (1989). Nonetheless, in these papers price discrimination does not occur at the firm level, but across firms. That is, each firm offers a single price in equilibrium, while the market distribution of prices effectively segments the consumer population into informed and uninformed buyers. Katz's (1984) model is an exception to these papers by introducing the ability of firms to set multiple prices to sort between informed and uninformed consumers; this contribution is reviewed in Section 3.5. At present, competitive analogs of Salop's (1977) monopoly price discrimination model, in which each firm offers multiple prices in equilibrium, have not been well explored.

tion policies,⁸ collusion and intertemporal price discrimination,⁹ price discrimination in aftermarkets¹⁰, advance-purchase discounts,¹¹ price discrimination in telecommunications,¹² and the strategic effect of product lines in imperfectly competitive settings.¹³

It is well known that price discrimination is only feasible under certain conditions: (i) firms have short-run market power, (ii) consumers can be segmented either directly or indirectly, and (iii) arbitrage across differently priced goods is infeasible. Given that these conditions are satisfied, an individual firm will typically have an incentive to price discriminate, holding the behavior of other firms constant. The form of price discrimination will depend importantly on the nature of market power, the form of consumer heterogeneity, and the availability of various segmenting mechanisms.

When a firm is a monopolist, it is simple to catalog the various forms of price discrimination according to the form of consumer segmentation. To this end, suppose that a consumer's preferences for a monopolist's product are given by

$$U = v(q, \theta) - y,$$

where q is the amount (quantity or quality) consumed of the monopolist's product, y is numeraire, and consumer heterogeneity is captured in $\theta = (\theta_o, \theta_u)$. The vector θ has two components. The first component, θ_o , is observable and prices may be conditioned upon it; the second component, θ_u , is unobservable and is known only to the consumer. We say that the monopolist is practicing *direct* price discrimination to the extent that its prices depend upon observable heterogeneity. Generally, this implies that the price of purchasing q units of output will be a function that depends upon θ_o : $P(q, \theta_o)$. When the firm further chooses to offer linear price schedules, $P(q, \theta_o) = p(\theta_o)q$, we say the firm is practicing *third-degree* price discrimination over the characteristic θ_o . If

⁸ While the chapter gives a flavor of a few of the influential papers on this topic, the treatment is left largely incomplete.

⁹ For instance, Gul (1987) shows that when durable-goods oligopolists can make frequent offers, unlike the monopolist, they improve their ability to commit to high prices and obtain close to full-commitment monopoly profits.

¹⁰ Tirole (1988, Section 3.3.1.5) provides a summary of the value of metering devices (such as two-part tariffs) as a form of price discrimination. Klein and Wiley (2003a) argue that such a form of price discrimination is common in competitive environments. In many regards, the economic intuition of two-part tariffs as a metering device is the same as the general intuition behind non-linear pricing which is surveyed in Section 6.

¹¹ Miravete (1996) and Courty and Li (2000) establish the economics behind sequential screening mechanisms such as advance-purchase discounts. Dana (1998) considers advance-purchase discounts in a variation of Prescott's (1975) model of perfect competition with aggregate demand uncertainty. Gale and Holmes (1992, 1993) provide a model of advance-purchase discounts in a duopoly setting.

¹² For example, see Laffont, Rey and Tirole (1998) and Dessein (2003).

¹³ A considerable amount of study has also focused on how product lines should be chosen to soften second-stage price competition. While the present survey considers the effect of product line choice in segmenting the marketplace (e.g., second-degree price discrimination), it is silent about the strategic effects of locking into a particular product line (i.e., a specific set of locations in a preference space). A now large set of research has been devoted to this important topic, including significant papers by Brander and Eaton (1984), Klempner (1992) and Gilbert and Matutes (1993), to list a few.

there is no unobservable heterogeneity and consumers have rectangular demand curves, then all consumer surplus is extracted and third-degree price discrimination is *perfect* price discrimination. More generally, if there is additional heterogeneity over θ_u or downward-sloping individual demand curves, third-degree price discrimination will leave some consumer surplus.

When the firm does not condition its price schedule on observable consumer characteristics, every consumer is offered the same price schedule, $P(q)$. Assuming that the unit cost is constant, we can say that a firm *indirectly* price discriminates if the marginal price varies across consumer types at their chosen consumption levels. A firm can typically extract greater consumer surplus by varying this price and screening consumers according to their revealed consumptions. This use of non-linear pricing as a sorting mechanism is typically referred to as *second-degree* price discrimination.¹⁴ More generally, in a richer setting with heterogeneity over both observable and unobservable characteristics, we expect that the monopolist will practice some combination of direct and indirect price discrimination – offering $P(q, \theta_o)$, while using the non-linearity of the price schedule to sort over unobservable characteristics.

While one can categorize price discrimination strategies as either direct or indirect, it is also useful to catalog strategies according to whether they discriminate across consumers (*interpersonal* price discrimination) or across units for the same consumer (*intrapersonal* price discrimination). For example, suppose that there is no interconsumer heterogeneity so that θ is fixed. There will generally remain some intraconsumer heterogeneity over the marginal value of each unit of consumption so that a firm cannot extract all consumer surplus using a linear price. Here, however, a firm can capture the consumer surplus associated with intraconsumer heterogeneity, either by offering a non-linear price schedule equal to the individual consumer's compensated demand curve or by offering a simpler two-part tariff. We will address these issues more in Section 5. Elsewhere in this chapter, we will focus on interpersonal price discrimination, with the implicit recognition that intrapersonal price discrimination often occurs in tandem.

The methodology of monopoly price discrimination can be both useful and misleading when applied to competitive settings. It is useful because the monopoly methods are valuable in calculating each firm's best response to its competitors' policies. Just as one can solve for the best-response function in a Cournot quantity game by deriving

¹⁴ Pigou (1920) introduced the terminology of first-, second- and third-degree price discrimination. There is some confusion, however, regarding Pigou's original definition of second-degree price discrimination and that of many recent writers [e.g., see Tirole (1988)], who include self-selection via non-linear pricing as a form of second-degree discrimination. Pigou (1920) did not consider second-degree price discrimination as a selection mechanism, but rather thought of it as an approximation of first-degree using a step function below the consumer's demand curve and, as such, regarded both first and second-degrees of price discrimination as "scarcely ever practicable" and "of academic interest only". Dupuit (1849) gives a much clearer acknowledgment of the importance of self-selection constraints in segmenting markets, although without a taxonomy of price discrimination. We follow the modern use of the phrase "second-degree price discrimination" to include indirect segmentation via non-linear pricing.

a residual demand curve and proceeding as if the firm was a monopolist on this residual market, we can also solve for best responses in more complex price discrimination games by deriving residual market demand curves. Unfortunately, our intuitions from the monopoly models can be misleading because we are ultimately interested in the equilibrium of the firms' best-response functions rather than a single optimal pricing strategy. For example, while it is certainly the case that, *ceteris paribus*, a single firm will weakly benefit by price discriminating, if every firm were to switch from uniform pricing to price discrimination, profits for the entire industry may fall.¹⁵ Industry profits fall depending on whether the additional surplus extraction allowed by price discrimination (the standard effect in the monopoly setting) exceeds any losses from a possibly increased intensity of price competition. This comparison, in turn, depends upon market details and the form of price discrimination, as will be explained. The pages that follow evaluate these interactions between price discrimination and imperfect competition.

When evaluating the impact of price discrimination in imperfectly competitive environments, two related comparisons are relevant for public policy. First, starting from a setting of imperfect competition and uniform pricing, what are the welfare changes from allowing firms to price discriminate? Second, starting from a setting of monopoly and price discrimination, what are the welfare effects of increasing competition? Because the theoretical predictions of the models often depend upon the nature of competition, consumer preferences and consumer heterogeneity, we shall examine a collection of specialized models to illuminate the broad themes of this literature and illustrate how the implications of competitive price discrimination compare to those of uniform pricing and monopoly.

In Sections 2–7, we explore variations on these themes by varying the forms of competition, preference heterogeneity, and segmenting devices. Initially in Section 2, we begin with the benchmark of first-degree, perfect price discrimination. In Section 3, we turn to the classic setting of third-degree price discrimination, applied to the case of imperfectly competitive firms. In Section 4, we examine an important class of models that extends third-degree price discrimination to dynamic settings, where price offers may be conditioned on a consumer's purchase history from rivals – a form of price discrimination that can only exist under competition. In Section 5, we study intrapersonal price discrimination. Section 6 brings together several diverse theoretical approaches to modeling imperfectly competitive second-degree price discrimination, comparing and contrasting the results to those under monopoly. Product bundling (as a form of price discrimination) in imperfectly competitive markets is reviewed in Section 7. Models of demand uncertainty and price rigidities are introduced in Section 8.

¹⁵ Even a monopolist may be worse off with the ability to price discriminate when commitment problems are present. For example, the Coase conjecture argues a monopolist is worse off with the ability to charge different prices over time. As another example, when a monopolist can offer secret price discounts to its retailers, the monopolist may end up effectively competing against itself in the retail market.

2. First-degree price discrimination

First-degree (or perfect) price discrimination – which arises when the seller can capture all consumer surplus by pricing each unit at precisely the consumer’s marginal willingness to pay – serves as an important benchmark and starting point for exploring more subtle forms of pricing. When the seller controls a monopoly, the monopolist obtains the entire social surplus, and so profit maximization is synonymous with maximizing social welfare. How does the economic intuition of this simple model translate to oligopoly?

The oligopoly game of perfect price discrimination is quite simple to analyze, even in its most general form. Following Spulber (1979), suppose that there are n firms, each selling a differentiated substitute product, but that each firm has the ability to price discriminate in the first degree and extract all of the consumer surplus under its residual demand curve. Specifically, suppose the residual demand curve for firm i is given by $p_i = D_i(q_i, q_{-i})$, when its rivals perfectly price discriminate and sell the vector $q_{-i} = (q_1, \dots, q_{i-1}, q_{i+1}, \dots, q_n)$. In addition, let $C_i(q_i)$ be firm i ’s cost of producing q_i units of output; the cost function is increasing and convex. As Spulber (1979) notes, the ability to then perfectly price discriminate when selling q_i units of output implies that firm i ’s profit function is

$$\pi_i(q_i, q_{-i}) = \int_0^{q_i} D_i(y, q_{-i}) dy - C_i(q_i).$$

A Nash equilibrium is a vector of outputs, (q_1^*, \dots, q_n^*) , such that each firm’s output, q_i^* , is a best-response to the output vector of its rivals, q_{-i}^* : formally, for all i and q_i , $\pi_i(q_i^*, q_{-i}^*) \geq \pi_i(q_i, q_{-i}^*)$. As Spulber (1979) notes, the assumption of perfect price discrimination – in tandem with the assumption that residual demand curves are downward sloping – implies that each firm i ’s profit function is strictly concave in its own output for any output vector of its rivals. Hence, the existence of a pure-strategy Nash equilibrium in quantities follows immediately. The equilibrium allocations are entirely determined by marginal-cost pricing based on each firm’s residual demand curve: $D_i(q_i^*, q_{-i}^*) = C_i'(q_i)$.¹⁶

In this equilibrium of perfect price discrimination, each consumer’s marginal purchase is priced at marginal cost, so, under mild technical assumptions, social surplus is maximized for a fixed number of firms.¹⁷ In this setting, unlike the imperfect price discrimination settings which follow, the welfare effect of price discrimination is immediate, just as with perfect price discrimination under monopoly. Note, however, that

¹⁶ This general existence result contrasts with the more restrictive assumptions required for pure-strategy Nash equilibria in which firms’ strategies are limited to choosing a fixed unit price for each good. Spulber (1979) also demonstrates that if an additional stability restriction on the derivatives of the residual demand curves is satisfied, this equilibrium is unique.

¹⁷ When firms can choose their product characteristics, as Spence (1976a) noted, if sellers can appropriately price discriminate, the social distortions are eliminated. Lederer and Hurter (1986) and MacLeod, Norman and Thisse (1988) more generally show that *for a fixed number of firms*, an equilibrium exists with efficient product choices. These papers assume that the industry configuration of firms is fixed.

while consumers obtain none of the surplus under the residual demand curves, it does not follow that consumers obtain no surplus at all; rather, for each firm i , they obtain no surplus from the addition of the i th firm's product to the current availability of $n - 1$ other goods. If the goods are close substitutes and marginal costs are constant, the residual demand curves are highly elastic and consumers may nonetheless obtain considerable surplus from the presence of competition. The net effect of price discrimination on total consumer surplus requires an explicit treatment of consumer demand. It may also be the case that each firm's residual demand curve is more elastic when its rivals can perfectly price discriminate than when they are forced to price uniformly. Thus, while each firm prefers the ability to perfectly price discriminate itself, the total industry profit may fall when price discrimination is allowed, depending on the form of competition and consumer preferences. We will see a clear example of this in the Hotelling-demand model of [Thisse and Vives \(1988\)](#) which examines discriminatory pricing based on observable location. In this simple setting, third-degree price discrimination is perfect, but firms are worse off and would prefer to commit collectively to uniform-pricing strategies.

The above conclusions take the number of firms, n , and their product characteristics as fixed. If the industry configuration is endogenous, additional welfare costs can arise. For example, if entry with fixed costs occurs until long-run profits are driven to zero, *and* if consumer surplus is entirely captured by price discrimination, then price discrimination lowers social welfare compared to uniform pricing. This conclusion follows immediately from the fact that when consumer surplus is zero and entry dissipates profits, there is zero social surplus generated by the market. Uniform pricing typically leaves some consumer surplus (and hence positive social welfare). In a general setting of spatial competition in which firms choose to enter and their product characteristics, [Bhaskar and To \(2004\)](#) demonstrate that perfect price discrimination always causes too much entry from a social welfare perspective. This is because the marginal firm captures its marginal social contribution relative to an inefficient allocation (rather than an efficient allocation of $n - 1$ firms), which over compensates entry. The intuition in [Bhaskar and To \(2004\)](#) is simply put: Suppose that the vector of product characteristics for the market is given by x , and let x_{-i} be the vector without the i th firm. Let $x^*(n)$ maximize social welfare when there are n firms supplying output, and let $W(x^*(n))$ be the associated social surplus. With $n - 1$ firms in the market, the analogous characteristics vector and welfare level are $x^*(n - 1)$ and $W(x^*(n - 1))$, respectively. Now suppose that firm i is the marginal firm. Then its profits under perfect price discrimination are equal to $W(x^*(n)) - W^*(x_{-i}^*(n))$. Note that the firm's social contribution is $W(x^*(n)) - W^*(x^*(n - 1))$. Because $W(x^*(n - 1)) > W(x_{-i}^*(n))$, incentives for entry are excessive.

Rather than further explore the stylized setting of first-degree price discrimination, the ensuing sections turn to explicit analyses undertaken for a variety of imperfect price discrimination strategies.

3. Third-degree price discrimination

The classic theory of third-degree price discrimination by a monopolist is straightforward: the optimal price-discriminating prices are found by applying the familiar inverse-elasticity rule to each market separately. If, however, oligopolists compete in each market, then each firm applies the inverse-elasticity rule using its own residual demand curve – an equilibrium construction. Here, the cross-price elasticities of demand play a central role in determining equilibrium prices and outputs. These cross-price elasticities, in turn, depend critically upon consumer preferences and the form of consumer heterogeneity.

3.1. Welfare analysis

In the static setting of third-degree price discrimination, there are three potential sources of social inefficiency. First, aggregate output over all market segments may be too low if prices exceed marginal cost. To this end, we seek to understand the conditions for which price discrimination leads to an increase or decrease in aggregate output relative to uniform pricing. Second, for a given level of aggregate consumption, price discrimination will typically generate interconsumer misallocations relative to uniform pricing; hence, aggregate output will not be efficiently distributed to the highest-value ends. Third, there may be cross-segment inefficiencies as a given consumer may be served by an inefficient firm, perhaps purchasing from a more distant or higher-cost firm to obtain a price discount. A fourth set of distortions related to the dynamics of entry is taken up in Section 3.5.

In the following, much is made of the relationship between aggregate output and welfare. If the same aggregate output is generated under uniform pricing as under price discrimination, then price discrimination must necessarily lower social welfare because output is allocated with multiple prices. With a uniform price, interconsumer misallocations are not possible. Therefore, placing production inefficiencies aside, an increase in aggregate output is a necessary condition for price discrimination to increase welfare.¹⁸ Robinson (1933) makes this point in the context of monopoly, but the economic logic applies more generally to imperfect competition, providing that the firms are equally efficient at production and the number of firms is fixed.¹⁹ Because we can easily make

¹⁸ Yoshida (2000), building on the papers of Katz (1987) and DeGraba (1990), considers third-degree price discrimination by an intermediate goods monopolist with Cournot competition in the downstream market, and finds a more subtle relationship between output and welfare. When demands and marginal costs are linear, price discrimination has no effect on the aggregate amount of intermediate good sold, but it may lead to an inefficient allocation of output among the downstream firms. Indeed, in Yoshida's (2000) model, if final output rises, social welfare must fall.

¹⁹ Schmalensee (1981) extends Robinson's (1933) analysis to more than two segments and provides a necessary condition for aggregate quantity to rise in this general setting. Varian (1985) generalizes Schmalensee's (1981) necessary condition by considering more general demand and cost settings, and also by establishing

statements only about aggregate output for many models of third-degree price discrimination, this result proves useful in giving limited power to draw welfare conclusions.

To this end, it is first helpful to review the setting of monopoly in Robinson (1933). That work concludes that when a monopolist price discriminates, whether aggregate output increases depends upon the relative curvature of the segmented demand curves.²⁰ Particularly in the case of two segments, if the “adjusted concavity” (an idea we will make precise below) of the more elastic market is greater than the adjusted concavity of the less elastic market at a uniform price, then output increases with price discrimination; when the reverse is true, aggregate output decreases. When a market segment has linear demand, the adjusted concavity is zero. It follows that when demand curves are linear – providing all markets are served – price discrimination has no effect on aggregate output.²¹ In sum, to make a determination about the effects of price discrimination on aggregate output under monopoly, one needs only to compare the adjusted concavities of each market segment. Below we will see that under competition, one will also need to consider the relationship between a firm’s elasticity and the market elasticity for each market segment.

There are two common approaches to modeling imperfect competition: quantity competition with homogeneous goods and price competition with product differentiation. The simple, quantity-competition model of oligopoly price discrimination is presented in Section 3.2. Quantity competition, however, is not a natural framework in which to discuss price discrimination, so we turn our focus to price-setting models of competition. Within price-setting games, we further distinguish two sets of models based upon consumer demands. In the first setting (Section 3.3), all firms agree in their ranking of high-price (or “strong”) markets and low-price (or “weak”) markets. Here, whether the strong market is more competitive than the weak market is critical for many economic conclusions. In the second setting (Section 3.4), firms are asymmetric in their ranking of strong and weak markets; e.g., firm *a*’s strong market is firm *b*’s weak market and conversely. With such asymmetry, equilibrium prices can move in patterns that are not possible under symmetric rankings and different economic insights present themselves. Following the treatment of price-setting games, we take up the topics of third-degree price discrimination with endogenous entry (Section 3.5) and private restrictions on price discrimination (Section 3.6).

a lower (sufficient condition) bound on welfare changes. Schwartz (1990) considers cases in which marginal costs are decreasing. Additional effects arising from heterogeneous firms alter the application of these welfare results to oligopoly, as noted by Galera (2003). It is possible, for example, that price discrimination may lead to more efficient production across firms with differing costs, offsetting the welfare loss from consumer misallocations, and thus leading to a net welfare gain without an increase in aggregate output. In this paper, we largely examine models of equally efficient firms.

²⁰ More precisely, Robinson defines the *adjusted concavity* of a segment demand curve to capture the precise notion of curvature necessary for the result. Schmalensee (1981) provides a deeper treatment which builds upon Robinson’s work.

²¹ This finding was first noted by Pigou (1920), and so Robinson’s analysis can be seen as a generalization to non-linear demand.

3.2. Cournot models of third-degree price discrimination

Perhaps the simplest model of imperfect competition and price discrimination is the immediate extension of Cournot's quantity-setting, homogeneous-good game to firms competing in distinct market segments. Suppose that there are m markets, $i = 1, \dots, m$, and n firms, $j = 1, \dots, n$, each of which produces at a constant marginal cost per unit. The timing of the output game is standard: each firm j simultaneously chooses its output levels for each of the i markets: $\{q_1^j, \dots, q_m^j\}$. Let the demand curve for each market i be given by $p_i = D_i(Q_i)$, where $Q_i = \sum_j q_i^j$, and suppose that in equilibrium all markets are active. Setting aside issues of equilibrium existence, it follows that the symmetric equilibrium outputs, $\{q_1^*, \dots, q_m^*\}$, satisfy, for every market i ,

$$MC = D_i(nq_i^*) + D_i'(nq_i^*)q_i^* = p_i^* \left(1 - \frac{1}{n\varepsilon_i^m}\right),$$

where ε_i^m is the market elasticity for segment i .

Several observations regarding the effects of competition follow immediately from this framework. First, marginal revenues are equal across market segments, just as in monopoly. Second, under mild assumptions related to the change in elasticities and marginal costs, as the number of firms increases, the markup over marginal cost decreases in each market segment. It follows that each firm's profit also decreases and consumer surplus increases as n increases. Third, if each market segment has a constant elasticity of demand, relative prices across segments are constant in n and, therefore, an increase in firms necessarily decreases absolute price dispersion. Finally, in the spirit of monopolistic competition, one can introduce a fixed cost of production and allow entry to drive long-run profits to zero, thereby making the size of the market endogenous. In such a setting, both long-term market power and economic profit are zero, but fixed costs of entry generate short-run market power, short-run economic rents, and prices above marginal cost.

Aside from the effects of competition, one can also inquire about the welfare effects of price discrimination relative to uniform pricing. For Cournot oligopoly, the adjusted concavities are key to determining the effect of price discrimination on aggregate output. When demand curves are linear, these concavities are zero, and – providing all market segments are served – price discrimination has no effect on total sales. To see this clearly, let $Q_i = \alpha_i - \beta_i p_i$ be the demand function for segment i (or alternatively, $p_i = D_i(Q_i) = \alpha_i/\beta_i - Q_i/\beta_i$), and therefore $Q = \alpha - \beta p$ is the aggregate output across all segments at a uniform price of p , where $\alpha \equiv \sum_i \alpha_i$ and $\beta \equiv \sum_i \beta_i$. With constant marginal cost of production, c , the Cournot–Nash equilibrium under uniform pricing (i.e., with one aggregated market) is simply $Q^u = (\alpha - \beta c) \left(\frac{n}{n+1}\right)$. Under price discrimination, the equilibrium total output in each segment is similarly given by $Q_i^{\text{pd}} = (\alpha_i - \beta_i c) \left(\frac{n}{n+1}\right)$. Summing across segments, aggregate output under price discrimination is equal to that under uniform pricing: $Q^{\text{pd}} = \sum_i Q_i^{\text{pd}} = Q^u$. Given that

firms are equally efficient at production and aggregate output is unchanged, price discrimination reduces welfare because it generates interconsumer misallocations. More generally, when demand curves are non-linear or some markets would not be served under uniform pricing, price discrimination may increase welfare. The ultimate conclusion is an empirical matter.

3.3. *A tale of two elasticities: best-response symmetry in price games*

In her study of third-degree price discrimination under monopoly, Robinson (1933) characterizes a monopolist's two markets as "strong" and "weak". By definition, a price discriminating monopolist sets the higher price in the strong market, and the lower price in the weak market. It is useful to extend this ranking to imperfectly competitive markets. Suppose that there are two markets, $i = 1, 2$. We say that market i is "weak" (and the other is "strong") for firm j if, for any uniform price(s) set by the other firm(s), the optimal price in market i is lower than the optimal price in the other market. Formally, if $BR_i^j(p)$ is the best-response function of firm j in market i , given that its rival sets the price p , then market 1 is weak (and 2 is strong) if and only if $BR_1^j(p) < BR_2^j(p)$ for all p . We say that the market environment satisfies *best-response symmetry* [using a phrase introduced by Corts (1998)] if the weak and strong markets of each firm coincide; alternatively, if the weak and strong markets of each firm differ, then the environment exhibits *best-response asymmetry*.²²

When firms agree on their rankings of markets from strong to weak, the introduction of price discrimination causes the price in the strong market to rise (and the price in the weak market to fall) relative to the uniform price. Whether aggregate output rises or falls depends upon the magnitude of these movements. In particular, there exists a useful result from Holmes (1989) which predicts when aggregate output will rise or fall with price discrimination, and therefore provides some indication about its ultimate welfare effects. This result is not available when best responses are asymmetric, which creates a crucial distinction in what follows. In this section, we assume that there exists best-response symmetry; in the following section, we study best-response asymmetry along the lines of Corts (1998).

Borenstein (1985) and Holmes (1989) extend the analysis of third-degree price discrimination to settings of imperfect competition with product differentiation, underscoring the significance of cross-price elasticities in predicting changes in profits and surplus. In particular, Holmes (1989) builds upon the monopoly model of Robinson (1933) and demonstrates that under symmetric duopoly, it is crucial to know the ratio of market to cross-price elasticities, aside from the adjusted concavities of demand. The curvatures of the market demand curves are insufficient, by themselves, to predict changes in aggregate output when markets are imperfectly competitive.

²² Note that a market might be "strong" for a monopoly but a "weak" market under duopoly if competition is more intense in the monopolist's "strong" market.

To understand the relevance of the ratio of market elasticity to cross-price elasticity, consider two markets, $i = 1, 2$, and duopolists, $j = a, b$, each offering products in both segments and producing with constant marginal cost of c per unit. We take market 2 to be the strong market and market 1 to be weak. Demand for firm j 's output in market i depends upon the prices offered by each firm in market i : $q_i^j(p_i^a, p_i^b)$. These demand functions are assumed to be symmetric across firms (i.e., symmetric to permuting indices a and b), so we can write $q_i(p) \equiv q_i^a(p, p) \equiv q_i^b(p, p)$. The *market elasticity* of demand in market i (as a function of symmetric price $p = p_i^a = p_i^b$) is therefore

$$\varepsilon_i^m(p) = -\frac{p}{q_i(p)}q_i'(p).$$

Furthermore, j 's own-price *firm elasticity* of demand in market i is

$$\varepsilon_{i,j}^f(p^a, p^b) = -\frac{p^j}{q_i^j(p^a, p^b)} \frac{\partial q_i^j(p^a, p^b)}{\partial p^j},$$

which at symmetric prices, $p = p_i^a = p_i^b$, is more simply

$$\varepsilon_i^f(p) = -\frac{p}{q_i(p)}q_i'(p) + \frac{p}{q_i(p)} \frac{\partial q_i^a(p, p)}{\partial p_i^b} = \varepsilon_i^m(p) + \varepsilon_i^c(p),$$

where $\varepsilon_i^c(p) > 0$ is the cross-price elasticity of demand at symmetric prices, p . Thus, the firm elasticity in a duopoly market is composed of two components: the market (or industry) elasticity and the cross-price elasticity. The former is related to the ability of a monopolist (or collusive duopoly) to extract consumer surplus; it measures the sensitivity of the consumer to taking the outside option of not consuming either good. The latter is related to the ability of a rival to steal business; it measures the consumer's sensitivity to purchasing the rival's product. While a monopolist will choose prices across markets such that

$$\frac{p_i - c}{p_i} = \frac{1}{\varepsilon_i^m(p_i)},$$

non-cooperative duopolists (in a symmetric price equilibrium) will set prices across markets such that

$$\frac{p_i - c}{p_i} = \frac{1}{\varepsilon_i^m(p_i) + \varepsilon_i^c(p_i)}.$$

Several results follow from this comparison.

- *Price effects of competition* From the above formulation of the inverse-elasticity rules, competition lowers prices in both markets compared to a price-discriminating monopolist, *ceteris paribus*, and therefore we expect competition to increase welfare in this simple third-degree price discrimination setting. The effect of competition on price dispersion across markets, however, is ambiguous and depends upon the cross-price elasticities. If the goods are close substitutes and market competition is fierce (i.e.,

$\varepsilon_1^c(p_1) \approx \infty$, $\varepsilon_2^c(p_2) \approx \infty$), prices will be close to marginal cost in each market and the price differential across markets will be negligible. Alternatively, if consumers in the weak market find the goods to be close substitutes (i.e., $\varepsilon_1^c(p_1) \approx \infty$) while consumers in the strong market exhibit powerful brand loyalties (i.e., $\varepsilon_2^c(p_2) \approx 0$), then the firms choose highly competitive prices in the weak market and close-to-monopoly prices in the strong market. Competition, in this setting, leads to greater price differentials across markets, relative to that of a price-discriminating monopolist. Testable implications for price dispersion are directly tied to estimates of the cross-price elasticities in each market.²³

• *Output (and welfare) effects of price discrimination* A recurring policy question in the price discrimination literature is whether to allow third-degree price discrimination or to enforce uniform pricing. A key ingredient to understanding this question in the context of imperfectly competitive markets is the impact of price discrimination on output.

Consider the marginal profit to firm j from a change in price in market i (starting from a point of price symmetry):

$$D\pi_i(p) \equiv q_i^a(p) + (p - c) \frac{\partial q_i^a(p, p)}{\partial p_i^a}.$$

We further assume that these marginal profit functions decrease in price for each market segment. The third-degree discriminatory prices are determined by the system of equations, $D\pi_i(p_i^*) = 0$, $i = 1, 2$, while the uniform-price equilibrium is determined by $D\pi_1(p_u^*) + D\pi_2(p_u^*) = 0$.²⁴ Given our assumption of decreasing marginal profit functions, it is necessarily the case that $p_u^* \in (p_1^*, p_2^*)$.²⁵ This in turn implies that the output in market 2 decreases under price discrimination while the output in market 1 increases. The impact of price discrimination on aggregate output, therefore, is not immediately clear.

To determine the effect on aggregate output, suppose that due to arbitrage difficulties, a discriminating firm cannot drive a wedge greater than r between its two prices²⁶; hence, $p_2 = p_1 + r$. It follows that, for a given binding constraint r , each firm will choose p_1 to satisfy $D\pi_1(p_1) + D\pi_2(p_1 + r) = 0$, the solution of which we denote

²³ Borenstein (1985) and Borenstein and Rose (1994) develop related theories indicating how competition may increase price dispersion. Borenstein and Rose (1994) find empirical evidence of various measures of increased price dispersion as a function of increased competition in airline ticket pricing.

²⁴ We assume that portions of each market are served under both forms of pricing.

²⁵ In the context of monopoly, the direction of price changes in third-degree price discrimination follows this pattern if the monopolist's profit function is strictly concave in price within each segment. When this is not the case (e.g., profit is bimodal in price), the direction of price changes is ambiguous, as shown by Nahata, Ostaszewski and Sahoo (1990).

²⁶ Leontief (1940) analyzes the effect of such a constraint on monopoly pricing. Schmalensee (1981) uses the technique to analyze aggregate output, as we do here.

parametrically as $p_1^*(r)$. By construction, $p_2^*(r) \equiv p_1^*(r) + r$. Hence, the effect on aggregate output from a fixed price differential of r can be characterized by

$$Q(r) = q_1(p_1^*(r)) + q_2(p_1^*(r) + r).$$

Because $r = 0$ corresponds to uniform pricing, it follows that if $Q(r)$ is everywhere increasing in r , then aggregate output increases from price discrimination. Alternatively, if $Q(r)$ is everywhere decreasing in r , aggregate output (and welfare) necessarily decreases. After some simplification, the condition that $Q'(r) > 0$ can be shown to be equivalent to the condition

$$\left[\frac{(p_2 - c)}{2q_2'(p_2)} \frac{d}{dp_2} \left(\frac{\partial q_2^a(p_2, p_2)}{\partial p_2^a} \right) - \frac{(p_1 - c)}{2q_1'(p_1)} \frac{d}{dp_1} \left(\frac{\partial q_1^a(p_1, p_1)}{\partial p_1^a} \right) \right] + \left[\frac{\varepsilon_2^c(p_2)}{\varepsilon_2^m(p_2)} - \frac{\varepsilon_1^c(p_1)}{\varepsilon_1^m(p_1)} \right] > 0.$$

The first bracketed expression is a straightforward variation of Robinson's adjusted-concavity condition found in the case of monopoly. When demands are linear (and adjusted concavities are zero), the elasticity-ratio test gives a sufficient condition for increased output:

$$\frac{\varepsilon_1^m(p_1)}{\varepsilon_2^m(p_2)} > \frac{\varepsilon_1^c(p_1)}{\varepsilon_2^c(p_2)}.$$

If the strong market (market 2) is more sensitive to competition in the sense that $\varepsilon_i^c/\varepsilon_i^m$ is larger, then price discrimination causes the strong market's output reduction to be less than the weak market's output increase. Accordingly, aggregate output rises. If the reduction in the strong market is sufficiently small relative to the weak market, then welfare will also rise. As an extreme example, when marginal cost is zero, optimal pricing requires that $\varepsilon_i^f(p_i) = 1$, and so $\varepsilon_i^m(p_i) = 1 - \varepsilon_i^c(p_i)$. The elasticity-ratio condition simplifies to the requirement that the strong market has a higher cross-price elasticity of demand: $\varepsilon_2^c(p_2) > \varepsilon_1^c(p_1)$.

From a social welfare point of view, the higher price should occur in the market with the lower market elasticity. This is also the pattern of monopoly pricing. Because each duopolist cares about its individual firm elasticity (which is the sum of the market and cross-price elasticities), this ordering may fail under competition and the higher price may arise in the more elastic market. While the average monopoly price exceeds the average duopoly price, the pattern of relative prices may be more inefficient under duopoly than under monopoly when firms are allowed to price discriminate. While competition is effective at controlling average prices, it is not effective at generating the correct pattern of relative prices. This is a key source of ambiguity in the welfare analysis.²⁷

²⁷ I am grateful to Mark Armstrong for providing this insight, which provides further intuition regarding the desirability of intrapersonal price discrimination in competitive markets (Section 5).

• *Profit effects* The profit effects of price discrimination are more difficult to predict. While any individual firm's profit rises when allowed to price discriminate, the entire industry profit may rise or fall when all firms add price discrimination to their strategic arsenals. Two papers have made significant findings in this direction. Holmes (1989) analyzes the case of linear demand functions and finds that when the elasticity-ratio condition above is satisfied, profit (as well as output) increases. When the condition is violated, however, the effect on profits is ambiguous (though welfare necessarily falls). What is particularly interesting is that price discrimination decreases profits when the weak market has a higher cross-price elasticity but a lower market elasticity compared to the strong market. Because the market elasticity is lower in the weak market, a given increase in price would be more profitable (and more socially efficient) in the weak market than in the strong market. When profits fall due to price discrimination, it is because the weak market's significantly higher cross-price elasticity outweighs its lower market elasticity, and therefore price discrimination reduces the weak-market price. From a profit perspective (and a social welfare perspective), this lower price is in the wrong market, and thus profits decline relative to uniform pricing.

Related to this finding, Armstrong and Vickers (2001) consider a model of third-degree price discrimination in which each segment is a Hotelling market with uniformly distributed consumers, each of whom has identical downward-sloping demand. The market segments differ only by the consumers' transportation costs. They demonstrate that when competition is sufficiently intense (specifically, each segment's transportation cost goes to zero while maintaining a constant cost ratio), industry profits increase under price discrimination and consumer surplus falls. This outcome suggests that Holmes's (1989) numerical examples of decreased profits (reductions which Holmes notes are never more than a few percentage points) may not be robust to closely-related settings with intense competition. This finding, together with the other results of Holmes (1989), strengthens the sense that price discrimination typically increases profits in settings of best-response symmetry with sufficient competition.²⁸

• *Inter-firm misallocations* The model examined in Holmes (1989) is symmetric across firms – no firm has a cost or product advantage over the other. This simplification obscures possibly significant, inter-firm misallocations that would arise in a model in which one firm has a comparative advantage in delivering value to consumers. The change from uniform pricing to price discrimination may either mitigate or amplify these distortions.

As an immediate illustration of this ambiguity, consider a duopoly setting in which both firms are local monopolists in their own strong markets and participate in a third, weak market. The strong-market demands are rectangular (although possibly different

²⁸ Armstrong and Vickers (2001) find that when the segment with the lower market elasticity also has a sufficiently higher cross-price elasticity (i.e., low transportation costs), welfare falls under price discrimination. The economics underlying this result are similar to Holmes's (1989) linear model of decreased profits – price discrimination causes the prices to fall and rise in the wrong markets.

between the firms). In the weak market, suppose that the firms are Hotelling (1929) duopolists in a market which is covered in equilibrium. Given these assumptions, the only social inefficiency is that some consumers in the weak market purchase from the “wrong” firm; this situation arises when the price differential across firms in the weak market is not equal to the difference in marginal costs. Under price discrimination, our assumption of rectangular demand curves implies that the strong-market, cross-firm price differential depends entirely on the consumer’s valuations for each product. In the weak market, however, it is easy to see that the resulting price differential between the firms is smaller than the difference in marginal costs; the price discrimination equilibrium results in the high-cost firm serving too much of the market. Compare this outcome to the uniform-price setting: If the strong market is sufficiently important relative to the weak market, then the uniform-price differential will be close to the price-discriminating, inter-firm differential in the strong market. If the strong-market differential is close to the difference in marginal costs, then uniform pricing mitigates inefficiencies; if the differential in the strong market is smaller than the price-discriminating differential in the weak market, then uniform pricing amplifies the social distortions. In short, there are no robust conclusions regarding the effect of price discrimination on misallocated production.

3.4. *When one firm’s strength is a rival’s weakness: best-response asymmetry in price games*

The assumption that firms rank strong and weak markets symmetrically is restrictive, as it rules out most models with spatial demand systems in which price discrimination occurs over observable location; e.g., a weak (distant) market for firm a is a strong (close) market for firm b . The assumption of best-response symmetry in the previous analysis allowed us to conclude that the uniform price always lies between the strong and weak market prices under discrimination. Without such symmetry, this conclusion does not generally hold. Indeed, it is possible that all prices rise or fall following the introduction of price discrimination.

- *A simple model to illustrate recurring themes* We begin with a simple example of differentiated duopoly drawn from Thisse and Vives (1988) to illustrate some of the consequences of price discrimination when firms have dissimilar strengths and weaknesses across markets.²⁹ Consider a standard Hotelling (1929) model of duopoly in which two firms are located on the endpoints of a linear market. Each consumer has an observable location parameter, $\theta \in (0, 1)$, which is drawn from a uniform distribution across the market; each consumer demands at most one unit of output. A consumer at location θ who consumes from the left firm at price p_l obtains utility $z - \tau\theta - p_l$, while

²⁹ See related papers by Shaffer and Zhang (1995, 2000), Bester and Petrakis (1996), Liu and Serfes (2004) and Armstrong (2006, Section 3.4).

the same consumer purchasing from the right firm at price p_r obtains $z - \tau(1 - \theta) - p_r$. In this sense, z represents the base value of the product, while τ is a measure of product differentiation and the intensity of competition. Each firm produces output at constant marginal (and average) cost equal to c .

In the analysis that immediately follows, we assume that z is sufficiently large so that the duopoly equilibrium exhibits competition (rather than local monopoly or kinked demand) and that the multi-plant monopolist covers the market. Critically, this assumption guarantees that industry demand is inelastic while the cross-price elasticity between products depends on $1/\tau$, which can be quite large. Among other things, the resulting relationship between the industry and cross-price elasticities induces intense competition in the duopoly setting.

As a benchmark, consider the case of uniform pricing. In the duopoly setting, it is straightforward to compute that in the Nash equilibrium, price is $p = c + \tau$ and each firm earns $\pi = \frac{1}{2}\tau$ in profits. Intuitively, a higher transportation cost translates into higher product differentiation, prices, and profits; indeed, τ is the unique source of profit in the Hotelling framework.

Now, suppose that firms are able to price discriminate *directly* on the consumer's location, θ , as in [Thisse and Vives \(1988\)](#). It follows in equilibrium that the more distant firm offers a price to the consumer of $p = c$ and the closer firm generates the same level of utility to make the sale.³⁰ Thus, the left firm offers a price of $p_l(\theta) = c + \tau(1 - 2\theta)$ for all $\theta \leq \frac{1}{2}$, and $p_l(\theta) = c$ for $\theta > \frac{1}{2}$; analogously, the right firm offers $p_r(\theta) = c + \tau(2\theta - 1)$ for $\theta \geq \frac{1}{2}$ and $p_r(\theta) = c$ for $\theta < \frac{1}{2}$. It immediately follows that price discrimination leads to a fall in equilibrium prices for *every* market segment: $p^d(\theta) \equiv \max\{p_l(\theta), p_r(\theta)\} < c + \tau$ for all $\theta \in (0, 1)$. Consequently, price discrimination also lowers profits which are now $\pi^d = \int_0^{\frac{1}{2}} \tau(1 - 2s) ds = \frac{1}{4}\tau$, – exactly half of the profits that arise under uniform pricing.

Compare these duopoly outcomes to those which emerge when the two products are sold by a multi-plant monopolist. When the monopolist is restricted to offering the output of each plant at a uniform mill price, the price will be set so as to leave the consumer at $\theta = \frac{1}{2}$ with no surplus (providing z is sufficiently large); hence, $p = z - \frac{1}{2}\tau$ and per-plant profits under uniform pricing are $\pi = \frac{1}{2}(z - c - \frac{1}{2}\tau)$. If the multi-plant monopolist is able to price discriminate over consumer location, however, it can do much better. The firm would offer a price to extract all consumer surplus, $p^m(\theta) = z - \tau \min\{\theta, 1 - \theta\}$, and per-plant profits would increase to $\pi^m = \frac{1}{2}(z - c - \frac{1}{4}\tau)$. Unlike imperfect competition, price discrimination increases the prices and profits of the monopolist.

Several noteworthy comparisons can be made. First, consider the effect of price discrimination on profits and price levels. Profits for a monopolist increase with the ability to price discriminate by $\tau/4$, while industry profits for competing duopolists decrease by

³⁰ [Lederer and Hurter \(1986\)](#) show that such marginal-cost pricing by the closest unsuccessful competitor is a common feature in pricing equilibria of location models.

$\tau/2$ when firms use discriminatory prices. Under duopoly, introducing price discrimination creates aggressive competition at every location and uniformly lowers every price: prices decrease from $p = c + \tau$ to $p^d(\theta) = c + \tau - 2\tau \min\{\theta, 1 - \theta\}$. Here, the business-stealing effect of price discrimination dominates the rent-extraction effect, so that duopolists are worse off with the ability to price discriminate. This result contrasts with that under best-response symmetry (Section 3.3) where price discrimination typically increases industry profits. Because welfare is constant in this simple setting, these profit conclusions imply that consumers are better off with price discrimination under competition and better off with uniform pricing under monopoly.³¹

Second, note that price discrimination generates a range of prices. Because perfect competition implies that all firms choose marginal cost pricing (even when allowed to price discriminate), a reasonable conjecture may be that an increase in competition reduces price dispersion relative to monopoly. In the present model, however, the reverse is true – competition increases dispersion.³² The range of prices with price discriminating duopolists is $p^d(\theta) \in [c, c + \tau]$, twice as large as the range for the multiplant monopolist, $p^m(\theta) \in [z - \frac{1}{2}\tau, z]$. Intuitively, when price discrimination is allowed, duopoly prices are more sensitive to transportation costs because the consumer's distance to the competitor's plant is critical. A multiplant monopolist, however, is only concerned with the consumer choosing the outside option of purchasing nothing, so the relevant distance for a consumer is that to the nearest plant – a shorter distance. More generally, monopoly prices are driven by market elasticities, while duopoly prices are determined in tandem with the cross-price elasticity with respect to a rival's price. This suggests that whether dispersion increases from competition depends upon the specifics of consumer preferences.³³

Third, if we generalize the model slightly so that firm a has higher unit costs than firm b , $c_a > c_b$, we can consider the impact of price discrimination on inter-firm misallocations. With uniform pricing, each firm chooses $p_j = \frac{2}{3}c_j + \frac{1}{3}c_{-j} + \tau$ and the price

³¹ When drawing welfare conclusions, one must be especially careful given this model's limitations. Because the market is covered in equilibrium and inelastic consumers purchase from the closest plant, it follows that all regimes (price discrimination or uniform pricing, monopoly or duopoly) generate the same total social surplus. Hence, a richer model will be required to generate plausible welfare conclusions; we will consider such models later. That said, this simple model is useful for generating immediate intuitions about the effects of competition and price discrimination on profit, consumer surplus, price levels and price dispersion. Most importantly, the model illustrates the significance of relaxing best-response symmetry over strong and weak markets.

³² Note that the range of duopoly prices contracts as τ becomes small. Because τ is a natural measure of the intensity of competition, we have the result that holding the number of firms fixed, increasing competitive pressures leads to less price dispersion.

³³ For example, if the base value also depends continuously upon location, $z(\theta)$, with $z'(\theta) < -\tau$ (respectively, $z'(\theta) > \tau$) for $\theta < \frac{1}{2}$ (respectively, $\theta > \frac{1}{2}$), then a multi-plant monopolist who sells to the entire market may offer a range of prices larger than would the duopolists. Whether competition increases or decreases price dispersion is unclear without more knowledge about market and cross-price elasticities. This ambiguity is no different from the setting of best-response symmetry; see Section 3.3.

differential is $(c_a - c_b)/3$. Because this price differential is smaller than the difference in firm costs, too many consumers purchase from the less efficient firm a . Under uniform pricing by a multi-plant monopolist, the differential is larger, $(c_a - c_b)/2$, but some misallocation remains. When price discrimination by location is allowed, however, efficiency in inter-firm allocations is restored for both duopoly and multi-plant monopoly settings. In the case of duopoly, efficiency arises because the marginal consumer, who is located at a point of indifference between the two endpoints, is always offered marginal cost pricing from both firms. In a monopoly, the monopolist extracts all consumer surplus using price discrimination and so eliminates any inter-plant misallocations. Although this setting is simplified, the intuition for why price discrimination decreases inter-firm misallocations can be generalized to a discrete-choice model in which the marginal consumer in equilibrium is offered marginal-cost prices under discrimination.

Fourth, consider the possibility of commitment. Given that profits are lower with price discrimination in the presence of competing firms, it is natural to wonder whether firms may find it individually optimal to commit publicly to uniform pricing in the hopes that this commitment will engender softer pricing responses from rivals. This question was first addressed in [Thisse and Vives \(1988\)](#), who add an initial commitment stage to the pricing game. In the first stage, each firm simultaneously decides whether to commit to uniform prices (or not); in the second stage, both firms observe any first stage commitments, and then each sets its price(s) accordingly.³⁴ One might conjecture that committing to uniform pricing may be a successful strategy if the second-stage pricing game exhibits strategic complementarities. It is straightforward to show in the present example, however, that the gain which a uniform-price firm would achieve by inducing a soft response from a rival is smaller than the loss the firm would suffer from the inability to price discriminate in the second stage. Formally, when one firm commits to uniform pricing and the other does not, the equilibrium uniform price is $p = c + \frac{1}{2}\tau$ and the optimal price-discriminatory response is $p(\theta) = c + \tau(\frac{3}{2} - 2\theta)$; these prices result in a market share of $\frac{1}{4}$ and a profit of $\pi = \frac{1}{8}\tau$ for the uniform pricing duopolist, and yield a market share of $\frac{3}{4}$ and a profit of $\pi = \frac{9}{16}\tau$ for the discriminator. Combined with the previous results on profits, it follows that choosing price discrimination in the first stage dominates uniform pricing: a Prisoners' Dilemma emerges.

- *General price effects from discrimination and competition* In the simple Hotelling model, prices decrease across all market segments when price discrimination is introduced. Because firms differ in their ranking of strong and weak markets in this example, one might wonder how closely this result depends upon best-response asymmetry. The answer, as [Corts \(1998\)](#) has shown, is that best-response asymmetry is a necessary condition for *all-out price competition* (defined as the case where prices drop in all markets

³⁴ When one firm commits to uniform pricing and the other does not, [Thisse and Vives \(1988\)](#) assume that the uniform-pricing firm moves first in the second-stage pricing game.

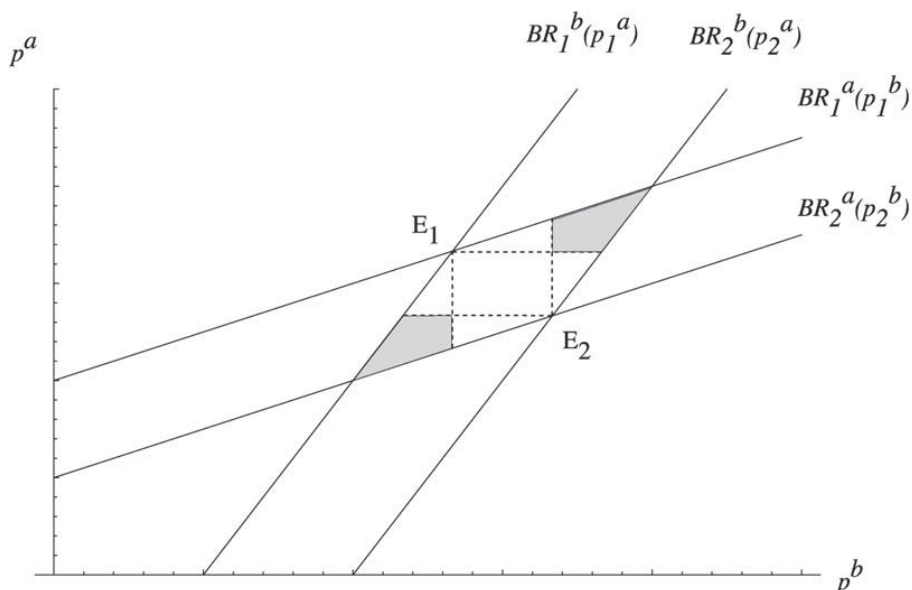


Figure 34.1. The possibility of all-out competition and all-out price gouging.

form competition) or all-out price increases (analogously, where prices increase in all markets).³⁵

Recall that if firms have identical rankings over the strength and weakness of markets (and if profit functions are appropriately concave and symmetric across firms), it follows that the uniform price lies between the weak and strong market prices. When firms do not rank markets symmetrically in this sense, it is no longer necessary that the uniform price lies between the price discriminating prices. Indeed, Corts (1998) demonstrates that for any profit functions consistent with best-response asymmetry, either all-out competition or all-out price increases may emerge from price discrimination, depending on the relative importance of the two markets.

Suppose there are two markets, $i = 1, 2$ and two firms, $j = a, b$, and that market 1 is firm a 's strong market but firm b 's weak market. Mathematically, this implies that the best-response functions satisfy the following inequalities: $BR_1^a(p) > BR_2^a(p)$ and $BR_1^b(p) < BR_2^b(p)$. Graphically, the price-discrimination equilibrium for each market is indicated by the intersections of the relevant best-response functions; these equilibria are labeled E_1 and E_2 in Figure 34.1.

Depending on the relative importance of market 1 and market 2, firm j 's uniform-price best-response function can be anywhere between its strong and weak best-response functions. With this insight, Corts (1998) shows that for any pair of prices

³⁵ Nevo and Wolfram (2002) and Besanko, Dube and Gupta (2003) empirically assess the possibility of all-out competition in breakfast cereals and ketchup, respectively. The former finds evidence consistent with best-response asymmetry and a general fall in prices in the breakfast cereal market; the latter, through empirical estimation of elasticities and simulations, finds no evidence of all-out competition in the branded ketchup market.

bounded by the four best-response functions, there exists a set of relative market weights (possibly different for each firm) that support these prices as an equilibrium. For example, if firm a finds market 1 to be sufficiently more important relative to market 2 and firm b has reverse views, then the uniform-price best-response functions intersect in the upper-right, shaded region, all-out competition emerges under price discrimination, and prices fall relative to the uniform-pricing regime. This is the intuition behind the simple Hotelling game above. There, each firm cares substantially more about its closer markets than its distant markets. On the other hand, if the uniform-price equilibrium is a pair of prices in the shaded region in the lower left, then price discrimination causes all segment prices to increase.

When the underlying demand functions induce either all-out competition or all-out price increases, the theoretical predictions are crisp. With all-out competition, price discrimination lowers all segment prices, raises all segment outputs, raises consumer utility and lowers firm profits. With all-out price increases, price discrimination has the opposite effects: all segment prices increase, all segment outputs decline, welfare and consumer surplus both decrease. When the underlying preferences do not generate all-out competition or price gouging (i.e., the uniform prices are not part of the shaded interiors of Figure 34.1) the impact of price discrimination is more difficult to assess. A more general treatment for settings of best-response asymmetry in which prices do not uniformly rise or fall would be useful to this end – perhaps focusing on market and cross-price elasticities in the manner of Holmes (1989).

3.5. Price discrimination and entry

The preceding analysis has largely taken the number of firms as exogenous. Given the possibility of entry with fixed costs, a new class of distortions arises: price discrimination may induce too much or too little entry relative to uniform pricing.

- *Monopolistic competition* If entry is unfettered and numerous potential entrants exist, entry occurs to the point where long-run profits are driven to zero. Under such models of monopolistic competition, social surplus is equated to consumer surplus. The question arises under free entry whether a change from uniform pricing to discrimination leads to higher or lower aggregate consumer surplus?

To answer this question, two effects must be resolved. First, by fixing the number of firms, does a change from uniform pricing to price discrimination lead to higher industry profits? We have already observed that price discrimination can either raise or lower industry profits, depending on the underlying system of demands. If price discrimination raises industry profits, then greater entry occurs; if price discrimination lowers profits, then fewer firms will operate in the market. Second, given that a move to price discrimination changes the size of the industry, will consumer surplus increase or decrease? Under uniform pricing, it is well known that the social and private values of entry may differ due to the effects of business stealing and product diversity; see, for example, Spence (1976b) and Mankiw and Whinston (1986). Generally, when comparing price

discrimination to uniform pricing, no clear welfare result about the social efficiency of free entry exists, although a few theoretical contributions are suggestive of the relative importance of various effects.

Katz (1984), in a model of monopolistic competition with price discrimination, was one of the first to study how production inefficiencies from excessive entry may arise. He found the impact of price discrimination on social welfare is ambiguous and depends upon various demand parameters.³⁶ Rather than developing Katz's (1984) model, this ambiguity can be illustrated with a few simple examples.

As a first example, consider varying the linear market in Thisse and Vives (1988) to a circular setting, as in Salop (1979), where inelastic unit-demand consumers are uniformly distributed around a circular market; as entry occurs, firms relocate equidistant from one another. When the market is covered, all potential consumers purchase a good from the nearest firm, so the optimal number of firms is that which minimizes the sum of transportation costs and the fixed costs of entry, K . The sum of costs is $nK + \tau/4n$ and the socially efficient level of entry is $n^{\text{eff}} = \frac{1}{2}\sqrt{\frac{\tau}{K}}$. Under uniform pricing, each firm chooses an equilibrium price of $p = c + \tau/n$, leading to per-firm profits of $\pi^u = \tau/n^2 - K$. Free entry implies that entry occurs until $\pi^u = 0$, so $n^u = \sqrt{\frac{\tau}{K}} > n^{\text{eff}}$. Twice the efficient level of entry occurs with uniform pricing; the marginal social cost of additional entry, K , exceeds the benefit of lower transportation costs that competition generates. In contrast, under price discrimination, rivals offer $\hat{p}(\theta) = c$ to consumers located more distant than $1/2n$; consumers purchase from the closest firm at a price of $p(\theta) = c + \tau(\frac{1}{n} - 2\theta)$. Equilibrium profits are lower under price discrimination for a given n , $\pi^{\text{pd}} = \frac{\tau}{2n^2} - K$, so entry occurs up to the point where $n^{\text{pd}} = \sqrt{\frac{\tau}{2K}}$. It follows that $n^u > n^{\text{pd}} > n^{\text{eff}}$, so price discrimination increases social welfare relative to uniform pricing by reducing the value of entry. This conclusion, of course, is limited to the model at hand.³⁷

It is also possible the price discrimination generates excessive entry relative to uniform pricing. To illustrate, suppose that consumer preferences are entirely observable so that a price-discriminating monopolist would capture all of the consumer surplus, unlike a uniform-pricing monopolist. To model imperfect competition, assume along the lines of Diamond (1971) that goods are homogeneous, but consumers must bear a small search cost for each visit to a non-local store to obtain a price quote. Each consumer obtains the price of the nearest (local) store at no cost. Then it is an equilibrium for all firms to offer the monopoly price schedule and for all consumers to purchase from their local store and not search. Firms follow identical pricing strategies, so there is no value to search, and each sells to its local consumers. Because of ex ante competition, firms enter this market until long-run profits are dissipated. Because more consumer surplus

³⁶ In particular, the proportion of informed to uninformed consumers is key.

³⁷ A similar result is found in Armstrong and Vickers (2001) but in the context of a different type of price discrimination.

remains under uniform pricing than under price discrimination, welfare is higher under uniform pricing while too much entry occurs under price discrimination.

- *Entry deterrence* Because it is natural to think that price discrimination may affect the ability of an incumbent firms to respond to local entry, entry deterrence and accommodation must also be considered in the welfare calculus.

One might imagine that firms are situated asymmetrically, as in [Armstrong and Vickers \(1993\)](#), where an incumbent firm serves two market segments, and potential entry can occur only in one of the segments. Here, price discrimination has no strategic value to the entrant given its limited access to a single market, but it does allow the incumbent to price lower in the newly entered market, while still maintaining monopoly profits from its captive segment. Hence, the incumbent's best-response discriminating prices following entry will generally result in a lower price in the attacked market than if uniform pricing across segments prevailed. It follows that the entrant's profits are higher following entry when price discrimination is banned. For sufficiently high entry costs, entry is blockaded whether or not the incumbent can price discriminate. For sufficiently low costs of entry, entry occurs regardless of whether the incumbent can price discriminate. For intermediate entry costs, the availability of price discrimination blockades entry, whereas uniform price restrictions accommodate the entrant and mitigate post-entry price reductions. Under uniform pricing, in [Armstrong and Vickers's \(1993\)](#) model, the prices in *both* markets are lower with entry than they would be with price discrimination and deterrence. Entrant profits and consumer surplus are also higher, while incumbent profits fall. This result is robust, providing that the monopoly price in the captive market exceeds the optimal discriminatory price in the competitive market. [Armstrong and Vickers \(1993\)](#) further demonstrate that the net welfare effects of uniform price restrictions are generally ambiguous, as the efficiencies from reduced prices must be offset against the inefficiencies from additional entry costs.³⁸

The model in [Armstrong and Vickers \(1993\)](#) illustrates the possibility that uniform pricing reduces all prices relative to price discrimination, due entirely to entry effects. A restriction to uniform prices promotes entry, which in turn generates the price reductions. A similar theme emerges in [Section 7](#) but for different economic reasons when we consider entry deterrence from bundling goods.

3.6. *Collective agreements to limit price discrimination*

There are two scenarios where we can make crisp comparisons of the social and collective incentives to price discriminate: (i) best-response asymmetry with all-out competition, and (ii) best-response symmetry with linear demands.

³⁸ See [Aguirre, Espinosa and Macho-Stadler \(1998\)](#) for extensions to this model which focus on the impact of other pricing policies, including FOB pricing, in this context. [Cheung and Wang \(1999\)](#) note that banning price discrimination may reduce entry when the monopoly price in the captive market is lower than the post-entry price of the competitive market.

When all-out competition is present, price discrimination lowers prices and profits. Hence, a collective agreement by firms to restrict price discrimination has the effect of raising prices for all consumers, lowering aggregate output, and lowering consumer surplus and total welfare. Collective incentives for discrimination are at odds with social welfare.

In the second setting of best-response symmetry with linear demands, Holmes (1989) demonstrates that if the elasticity-ratio condition is satisfied, then price discrimination increases industry profits. It follows that with linear demands, firms have a collective incentive to prohibit price discrimination only if the elasticity-ratio condition fails. Given that demands are linear, violation of the elasticity-ratio condition also implies that price discrimination lowers aggregate output and hence welfare. In this case, it follows that society benefits by allowing firms to agree collectively to limit price discrimination. There is also the possibility that the elasticity-ratio condition fails and price discrimination is welfare-reducing, but industry profits nonetheless increase under price discrimination. Here, restrictions on welfare-reducing price discrimination must come from outside the industry. No collective agreement restricting price discrimination would arise even though it is socially optimal. In short, with linear demands, a collective of firms would inadequately restrict price discrimination from a social vantage point.

Winter (1997) considers a variation of Holmes's (1989) analysis in the context of collective agreements to limit (but not prohibit) price discrimination by restricting the difference between the high and low prices. His conclusion for linear demands is similar: when firms have a collective desire to restrict price discrimination, it is socially efficient for them to be allowed to do so. As an illustration, suppose an extreme case in which each half of the strong market is captive to one of the firms (therefore, $\varepsilon_2^c = 0$), while the weak market has a positive cross-price elasticity of demand. In such a case, the elasticity-ratio test fails. At the equilibrium prices, a slight restriction on price discrimination causes the weak-market price to rise slightly and the strong-market price to fall by a similar margin. Because the weak market price is below the collusive profit-maximizing price, this price increase generates greater profits. Because the strong market's price is at the optimal monopoly price under discrimination (due to captive customers), a slight decrease causes only a negligible (second-order) reduction in profits. Hence, a slight restriction on price discrimination is jointly optimal for duopolists. Moreover, as a consequence of Holmes (1989), a restriction on the price differential (a lowering of r) raises aggregate output. Since aggregate output increases and the price differential decreases, welfare necessarily increases. It follows that industry agreements to limit price discrimination arise only if price discrimination reduces welfare, providing that adjusted demand concavities are insignificant. The results are less clear when demands are not linear and the adjusted concavity condition plays an important role.

In short, when profits are lower under price discrimination, firms in the industry would prefer to collude and commit to uniform pricing. Such collusion would decrease welfare if all-out competition would otherwise occur, and increase welfare when demand is linear and price discrimination would have reduced aggregate output.

• *Vertical restraints and downstream price discrimination* Although this survey largely ignores the impact of price discrimination on competing vertical structures, it is worth mentioning a sample of work in this area. We have already mentioned one strategic effect of price discrimination via secret price discounts by wholesalers to downstream firms: price discrimination may induce the upstream firm to flood the downstream market. A legal requirement that the wholesaler offer a single price to all retailers may instead induce the upstream firm to not over supply the retail market, thereby raising profits and retail prices.³⁹

In other settings of wholesale price discrimination, if downstream market segments have different elasticities of demand but third-degree price discrimination is illegal or otherwise impractical because of arbitrage, vertical integration can be used as a substitute for price discrimination. Tirole (1988, Section 4.6.2) gives a simple model of such a vertical price squeeze. A monopoly wholesaler, selling to a strong market at price p_2 and to a weak market at price $p_1 < p_2$, may suffer arbitrage as the firms in the weak downstream market resell output to the strong segment. By vertically integrating into one of the weak-segment downstream firms, the wholesaler can now supply all output at the strong-segment price of p_2 while producing in the weak segment and using an internal transfer price no greater than p_1 . Other firms in the weak market will be squeezed by the vertically integrated rival due to higher wholesale prices. The wholesaler effectively reduces competition in the weak segment to prevent arbitrage and to implement a uniform wholesale price.

Consider instead the case where it is the downstream retail firms that are the source of price discrimination. How do the various tools of resale price maintenance (RPM) by the upstream manufacturer impact profits and welfare when retailers engage in third-degree price discrimination? As the previous discussions suggested, a manufacturer who sells to imperfectly competitive, price-discriminating retailers would prefer to constrain retailers from discounting their prices to consumers who are highly cross-elastic, as this generates an unprofitable business-stealing externality. On the other hand, the manufacturer would like to encourage price discrimination across the full range of cross-price inelastic consumers as this action raises profits to the industry. Hence, the combination of competition and price discrimination generates a conflict in the vertical chain.

A simple duopoly example from Chen (1999) illustrates this conflict. Suppose that type-1 consumers are captive and buy only from the local retailer (if at all), while type-2 consumers comparison-shop for the lowest price. Both types have unit demands with reservation prices drawn from a uniform distribution on $[0, 1]$. Here, price discrimination arises simply because a consumer's outside option may depend upon his type (and the competing offer of a rival). It is assumed that competing retailers can identify consumer types costlessly, perhaps because only comparison-shopping consumers find coupons with targeted price reductions. Market 1 (comprised of type-1 consumers) has a measure of α ; market 2 has a measure of $(1 - \alpha)$. Marginal costs

³⁹ For more on this and related points, see Rey and Tirole (2007), elsewhere in this volume.

are 0. The optimal prices which maximize the sum of retailers' and manufacturer's profits are $p_1 = p_2 = \frac{1}{2}$; total profit is $\frac{1}{4}$. When $\alpha \in (0, 1)$, this collective maximum cannot be achieved with two-part tariffs by themselves. A two-part tariff of the form $T(q) = F + wq$ will generate equilibrium prices by the retailers of $p_1 = \frac{1+w}{2}$ and $p_2 = w$; the conditionally optimal fixed fee will extract retailer profits, $F = \frac{1}{4}\alpha(1-w)^2$. The optimal wholesale price in this setting is

$$w^* = \frac{2(1-\alpha)}{4-3\alpha},$$

which implies retail prices will exceed the profit-maximizing prices of $\frac{1}{2}$. In effect, a classic double marginalization arises on each market segment. With the addition of either price ceilings or price floors, the two-part tariff again becomes sufficient to maximize the vertical chain's profit. For example, either $w = \frac{1}{2}$ and $F = 0$ in tandem with a price ceiling of $\frac{1}{2}$, or $w = 0$ and $F = \frac{1}{4}$ in tandem with a price floor of $\frac{1}{2}$, will achieve the desired retail prices. Moreover, RPM here has the desirable effect of lowering prices, raising output and making prices less dispersed across markets. With more general demand settings (specifically, type-1 consumers' valuations distributed differently than type-2 consumers), RPM can again implement the vertical chain's optimal retail prices, but its welfare effects are ambiguous. [Chen \(1999\)](#) places bounds on welfare changes which provide, among other things, that if output increases due to RPM, then welfare is necessarily higher, but, if output decreases, the change in welfare is ambiguous.

4. Price discrimination by purchase history

Consumer price sensitivities are often revealed by past purchase decisions. There are two related cases to consider. First, consumers may suffer costs if they switch to new products, so past customers may have more inelastic demands for their chosen brand than new customers. For example, a buyer of a particular word processing package may have to expend considerable cost to learn how to use the product effectively. Having expended this cost, choosing a competing software program is less attractive than it would have been initially. Here, purchase history is useful because an otherwise homogeneous good becomes differentiated ex post due to exogenous switching costs. It follows that competing firms may have an incentive to pay consumers to switch. Second, it may be that no exogenous switching cost exists but that the products are inherently differentiated, with consumers having strong preferences for one product or another. It follows that a customer who reveals a preference for firm a 's product at current prices is precisely the person to whom firm b would like to offer a price reduction. In this case, purchase history operates through a different conduit of differentiation because it informs about a consumer's exogenous brand preference. Competing firms may have an incentive to lower prices selectively to poach consumers who purchased from a rival in the past. For example, consider a consumer who prefers a particular brand of frozen

food. Rival brands may wish to target price reductions to these consumers by contracting with grocery stores to selectively print coupons on the reverse side of their sales receipts.

Regardless, the strategies of “paying customers to switch” [as in Chen (1997b)] or “consumer poaching” [as in Fudenberg and Tirole (2000)] can be profitable because purchase history provides a valuable instrument for the basis of dynamic third-degree price discrimination. It is not surprising, therefore, that such behavior-based price discrimination is a well-known strategy in marketing science.⁴⁰

As the examples suggest, the literature takes two approaches to modeling imperfect competition and purchase-history price discrimination. The first set of models, e.g., Nilssen (1992), Chen (1997b), Taylor (2003), and others, assumes that the goods are initially homogeneous in period 1, but after purchase the consumers are partially locked in with their sellers; exogenous costs must be incurred to switch to different firms in future periods. The immediate result is that although prices rise over time as firms exploit the lock-in effects of switching costs, in period 1 firms compete over the future lock-in rents.

The second set of models assumes that products are horizontally differentiated in the initial period [e.g., Caminal and Matutes (1990), Villas-Boas (1999), Fudenberg and Tirole (2000), and others]. In the simplest variant, brand preferences are constant over time. It follows that a consumer who prefers firm *a*'s product and reveals this preference through his purchase in period 1 will become identified as part of *a*'s “strong” market segment (and firm *b*'s “weak” market segment) in period 2. As we will see, when firms cannot commit to long-term prices, this form of unchanging brand preference will generate prices that *decrease* over time, and competition intensifies in each market.

The resulting price paths in the above settings rest on the assumption that firms cannot commit to future prices. This assumption may be inappropriate. One could easily imagine that firms commit in advance to reward loyal customers in the future with price reductions or other benefits (such as with frequent flyer programs). Such long-term commitments can be thought of as endogenous switching costs and have been studied in the context of horizontal differentiation by Banerjee and Summers (1987), Caminal and Matutes (1990), Villas-Boas (1999) and Fudenberg and Tirole (2000). Two cases have been studied: preferences that change over time and preferences that are static. In the first case, most papers assume that consumer valuations are independently distributed across the periods. If preferences change from period to period and firms cannot commit to future prices, there is no value to using purchase history as it is uninformative

⁴⁰ Kotler (1994, ch. 11) refers to segmenting markets based upon purchasing history as “behavioral segmentation” (e.g., user status, loyalty status, etc.), as opposed to geographic, demographic or psychographic segmentation. Rossi and Allenby (1993) argue that firms should offer price reductions to households “that show loyalty toward other brands and yet are price sensitive” (p. 178). Rossi, McCulloch and Allenby (1996) review some of the available purchase-history data. Acquisti and Varian (2005) study a general model of pricing conditional on purchase history and its application in marketing. See Shaffer and Zhang (2000) for additional references to the marketing literature. See also the forthcoming survey by Fudenberg and Villas-Boas (2005) for a review of the issues presented in this section.

about current elasticities. Public, long-term contracts between, say, firm a and a consumer, however, can raise the joint surplus of the pair by making firm b price lower in the second period to induce switching, as in the models of Caminal and Matutes (1990) and Fudenberg and Tirole (2000). In equilibrium, social welfare may decrease as long-term contracts induce too little switching. In the second category of price-commitment models, preferences are assumed to be fixed across periods. When preferences are unchanging across time, Fudenberg and Tirole (2000) demonstrate a similar effect from long-term contracts: firm a locks in some of its customer base to generate lower second-period pricing for switching, and thereby encourages consumers to buy from firm a in the initial period. With long-term contracts, some inefficient switching still occurs but less than when firms cannot commit to long-term prices; hence welfare increases by allowing such contracts.

We consider both switching-cost and horizontal-differentiation models of pricing without commitment in the following two subsections. We then turn to the effects of long-term price commitments when discrimination on purchase history is allowed.

4.1. Exogenous switching costs and homogeneous goods

Farrell and Klemperer (2007), in this volume, provide a thorough treatment of switching costs, so we limit our present attention to the specific issues of price discrimination over purchase history under imperfect competition.

One of the first discussions of purchase-history discrimination in a model of switching costs appears in Nilssen (1992).⁴¹ Chen (1997b), builds on this approach by distributing switching costs in a two-period model, which results in some equilibrium switching. We present a variant of Chen's (1997b) model here.

Consider duopolists, $j = a, b$, selling identical homogeneous goods to a unit measure of potential consumers. In period 1, both firms offer first-period prices, p_1^j . Each consumer chooses a single firm from which to purchase one unit and obtains first-period utility of $v - p_1^j$. Consumers who are indifferent randomize between firms. Following the first-period purchase, each consumer randomly draws an unobservable switching cost, θ , that is distributed uniformly on $[0, \bar{\theta}]$. When price discrimination is allowed, firms simultaneously offer pairs of second-period prices, $\{p_{2a}^j, p_{2b}^j\}$, where p_{2k}^j is the second-period price offered by firm j to a consumer who purchased from firm k in period 1. A consumer who purchases from firm j in both periods obtains a present-value utility of

$$v - p_1^j + \delta(v - p_{2j}^j),$$

⁴¹ In this model, however, there is no uncertainty over the size of the switching cost, and no consumer actually switches in equilibrium. The focus in Nilssen (1992) is primarily on how market outcomes are affected by the form of switching costs. Transaction costs are paid every time the consumer switches in contrast to learning costs which are only paid the first time the consumer uses a firm's product. These costs are indistinguishable in the two-period models considered in this survey.

and, if the consumer switches from firm j to firm k with switching cost θ , obtains

$$v - p_1^j + \delta(v - p_{2j}^k - \theta).$$

Beginning with the second period, suppose that firm a acquired a fraction ϕ^a of consumers in the first period. It follows that a consumer who purchased from firm a is indifferent to switching to firm b if, and only if,

$$v - p_{2a}^a = v - p_{2a}^b - \theta.$$

Consequently, firm a 's retained demand is

$$\phi^a \int_{p_{2a}^a - p_{2a}^b}^{\bar{\theta}} \frac{1}{\bar{\theta}} d\theta = \phi^a \left(1 - \frac{p_{2a}^a - p_{2a}^b}{\bar{\theta}} \right),$$

and firm b 's switching demand is $\phi^a (p_{2a}^a - p_{2a}^b) / \bar{\theta}$, provided that the market is covered in equilibrium. It is straightforward to calculate the other second-period demand functions and profits as a function of second-period prices. Solving for the equilibrium, second-period, price-discriminating prices, we obtain $p_{2j}^j = c + \frac{2}{3}\bar{\theta}$ and $p_{2k}^j = c + \frac{1}{3}\bar{\theta}$ for $k \neq j$.⁴² Using these prices, the associated equilibrium second-period profits are

$$\pi^j = \frac{\bar{\theta}}{3} \left(\frac{1}{3} + \phi^j \right),$$

a function solely of first-period market share.

First-period competition is perfect as the firm's goods are homogeneous. Chen (1997b) demonstrates that the unique subgame perfect equilibrium is for each firm to charge $p_1^j = c - \frac{\delta}{3}\bar{\theta}$, generating a present value of profits, $\frac{\delta}{9}\bar{\theta}$. In equilibrium, each firm sets its price so that the additional market share generated by a price decrease exactly equals the discounted marginal profit, $\delta\bar{\theta}/3$, derived from above. No additional profit is made from acquired first-period market share, but the firms continue to earn positive long-term profits because of the ability to induce switching in the second period and the underlying heterogeneity in preferences. Indeed, a firm who acquires zero first-period market share still earns profits of $\frac{\delta}{9}\bar{\theta}$. Prices increase over time, but more so for those consumers who are locked in. In equilibrium, all consumer types with switching costs below $\frac{\bar{\theta}}{3}$ inefficiently switch firms (i.e., one-third of the consumers switch), thereby revealing their price sensitivity and obtaining the discounted price.

Compare the price-discrimination outcome with what would emerge under uniform-pricing restrictions. In the second period, equilibrium prices will generally depend upon first-period market shares (particularly, prices depend upon whether the market share is above, below, or equal to $\frac{1}{2}$). A firm with a higher market share will charge a higher

⁴² Remarkably, the second-period prices are independent of first-period market share, a result that Chen (1997b) generalizes to a richer model than presented here.

second-period price. When market shares are equal, computations reveal that second-period equilibrium prices are $p_2^j = c + \bar{\theta}$, $j = a, b$.⁴³ Hence, we are in the setting of all-out-competition as prices for both segments are lower under price discrimination relative to uniform pricing, consumer surpluses are higher, and firm profits are lower.

In the first period, unfortunately, the analysis is more complex because the second-period profit functions are kinked at the point where market shares are equal. This gives rise to multiple equilibria in which consumers realize that firms with larger market shares will have higher prices in the second period, and they take this into account when purchasing in period 1, leading to less elastic first-period demand. One appealing equilibrium to consider is a symmetric equilibrium in which the first-period prices constitute an equilibrium in a one-shot game.⁴⁴ Here, the equilibrium outcome is $p_1^j = c + \frac{2}{3}\bar{\theta}$, which is higher than the first-period price-discriminating price. Furthermore, the discounted sum of equilibrium profits is $\frac{5\delta}{6}\bar{\theta}$, which is higher than the price-discriminating level of $\frac{\delta}{9}\bar{\theta}$. Generally, [Chen \(1997b\)](#) demonstrates that regardless of the selected uniform-pricing equilibrium, the discounted sum of profits under uniform pricing is always weakly greater than the profits under price discrimination. Thus, price discrimination unambiguously makes firms worse off in the switching-cost model.

Consumers, on the other hand, may or may not be better off under uniform pricing, depending on the equilibrium chosen. In the selected “one-shot” equilibrium above, consumers are unambiguously better off under price discrimination. In other equilibria derived in [Chen \(1997b\)](#), both firms and consumers may be worse off under price discrimination. Regardless, price discrimination always reduces welfare because it induces inefficient switching, unlike uniform pricing. Of course, one must be careful when interpreting the welfare results in such models of inelastic demand as price discrimination has no role to increase aggregate output.

It might seem odd that firms earn positive profits in [Chen \(1997b\)](#) given that, *ex ante*, the duopolists’ goods are homogeneous and competition is perfect in the first period. Profits are earned because only one firm (a monopoly) induces switching in the second period. [Taylor \(2003\)](#) makes this point (among others) by noting that in a model with three firms over two periods, both outside firms perfectly compete over prices in the second period, leading to zero profits from switching consumers, $p_{2k}^j = c$ for $k \neq j$.⁴⁵

⁴³ Note that the computations for the uniform-price equilibrium are not straightforward because there is a kink at $\phi^a = \frac{1}{2}$ and profit functions are not differentiable at equilibrium prices. See [Chen \(1997b\)](#) for details.

⁴⁴ See [Chen \(1997b\)](#) for details.

⁴⁵ [Taylor \(2003\)](#) considers a more general T -period model; we simplify the present discussion by focusing on the 2-period variation. Taylor also studies a more complex setting of screening on unobservable characteristics. Specifically, Taylor assumes that there are two first-order, stochastically ranked distributions of switching costs. Consumers who draw from the low-cost distribution signal their type in equilibrium by switching, and hence generate lower prices in the future. This idea is closely related to the topic of second-degree price discrimination studied in Section 6 below.

At such prices, the inside firm will retain its consumer if and only if $\theta \geq p_{2j}^j - c$, so it chooses its retention price to maximize $(\bar{\theta} - (p_{2j}^j - c))(p_{2j}^j - c)$, or $p_{2j}^j = c + \frac{1}{2}\bar{\theta}$. Comparing this second-period price spread of $\frac{1}{2}\bar{\theta}$ to the previous duopoly spread of $\frac{1}{3}\bar{\theta}$, increased competition (going from duopoly to oligopoly, $n > 3$) leads to more inefficient switching and greater price dispersion. In a sense, this increased price dispersion is similar to the effect present in [Thisse and Vives \(1988\)](#) when one goes from monopoly to duopoly: increases in competition can differentially affect some market segments more than others, leading to a larger range of equilibrium prices. Here, going from two to three firms leads to increased competition among firms that induce switching, but does not influence the loyal customer price to the same degree.

[Shaffer and Zhang \(2000\)](#) consider a model similar to [Chen \(1997b\)](#), studying the case in which firms' demands are asymmetric and the effect these asymmetries may have on whether a firm "pays customers to switch" or instead "pays consumers to stay". It is, of course, always optimal for price-discriminating firms to offer a lower price to its more elastic segment of consumers. Focusing on the second period of a duopoly model, [Shaffer and Zhang \(2000\)](#) demonstrate that although charging a lower price to a competitor's customers is always optimal when demand is symmetric, with asymmetries it may be that one firm's more elastic consumer segment is its own customers. This latter situation arises, for example, when firm a 's existing customer base has lower average switching costs, compared to firm b 's loyal consumers. In such a setting, firm a finds its loyal consumer segment has a higher elasticity of demand than the potential switchers from firm b ; firm a will charge a lower price to its loyal segment as a result. Hence, defending one's consumer base with "pay-to-stay" strategies can be optimal in a more general model.

4.2. Discrimination based on revealed first-period preferences

Instead of assuming exogenous switching costs arise after an initial purchase from either firm, one may suppose that consumers have exogenous preferences for brands that are present from the start, as in [Fudenberg and Tirole \(2000\)](#). To simplify the analysis, they model such horizontal differentiation by imagining a Hotelling-style linear market of unit length with firms positioned at the endpoints, and by assuming that each consumer's uniformly distributed brand preference, θ , remains fixed for both periods of consumption. Consumers have transportation costs of τ per unit distance, and firms produce with constant marginal and average costs of c per unit. In such a setting, consumers reveal information about their brand preference by their first-period choice, and firms set second-period prices accordingly.

Solving backwards from the second period, suppose that firm a captures the market share $[0, \hat{\theta}_1)$ and firm b captures the complement, $(\hat{\theta}_1, 1]$ in the first period. The second-period demand function derivations are straightforward. In the left segment (i.e., firm a 's strong market and firm b 's weak market), the marginal consumer, $\hat{\theta}_2^a$, who is indifferent between continuing to purchase from firm a at price p_{2a}^a and switching to firm b at price

of p_{2a}^b , is given by

$$p_{2a}^a + \tau\theta_2^a = p_{2a}^b + \tau(1 - \theta_2^a).$$

It follows that firm a 's demand from retained consumers is

$$\hat{\theta}_2^a = \frac{1}{2} + \left(\frac{p_{2a}^b - p_{2a}^a}{2\tau} \right),$$

and firm b 's demand from switching consumers is

$$\hat{\theta}_2^b = \hat{\theta}_1 - \hat{\theta}_2^a = \hat{\theta}_1 - q_{2a}^a = \hat{\theta}_1 - \frac{1}{2} + \left(\frac{p_{2a}^a - p_{2a}^b}{2\tau} \right).$$

Similar derivations provide the retained demand for firm b and the switching demand for firm a . Using these demand functions, simple computations reveal that equilibrium second-period prices are $p_{2a}^a(\hat{\theta}_1) = p_{2b}^b(\hat{\theta}_1) = c + \frac{\tau}{3}(1 + 2\hat{\theta}_1)$ and $p_{2a}^b(\hat{\theta}_1) = p_{2b}^a(\hat{\theta}_1) = c + \frac{\tau}{3}(4\hat{\theta}_1 - 1)$. When the first-period market is equally split, we have $p_{2a}^a = p_{2b}^b = c + \frac{2}{3}\tau$ and $p_{2a}^b = p_{2b}^a = c + \frac{1}{3}\tau$.

The marginal consumer in the first period will ultimately switch in the second period, so the location $\hat{\theta}_1$ is determined by the relationship

$$p_1^a + \tau\hat{\theta}_1 + \delta(p_{2a}^b(\hat{\theta}_1) + \tau(1 - \hat{\theta}_1)) = p_1^b + \tau(1 - \hat{\theta}_1) + \delta(p_{2b}^a(\hat{\theta}_1) + \tau\hat{\theta}_1).$$

Simplifying, first-period demand is

$$\hat{\theta}_1(p_1^a, p_1^b) = \frac{1}{2} + \frac{3}{2\tau(3 + \delta)}(p_1^b - p_1^a).$$

Providing $\delta > 0$, first-period demands are *less* sensitive to prices relative to the static one-shot game because an increase in first-period market share implies higher second-period prices. Using the equilibrium prices from the second period as a function of $\hat{\theta}_1$, one can compute second-period market shares as a function of $\hat{\theta}_1$. Because $\hat{\theta}_1$ is a function of first-period prices, second-period market shares and prices are entirely determined by first-period prices: $\hat{\theta}_2^a(\hat{\theta}_1(p_1^a, p_1^b))$ and $\hat{\theta}_2^b(\hat{\theta}_1(p_1^a, p_1^b))$. With these expressions, the present value of profit for firm a can now be written as a function of first-period prices. Computing the equilibrium is algebraically tedious but straightforward, leading one to conclude that first-period prices are $p_1^a = p_1^b = c + \tau + \frac{\delta}{3}\tau$, second-period prices are $p_{2a}^a = p_{2b}^b = c + \frac{2}{3}\tau$ for loyal customers, and $p_{2a}^b = p_{2b}^a = c + \frac{1}{3}\tau$ for switchers. The first-period market is split symmetrically with $\hat{\theta}_1 = \frac{1}{2}$, and the second-period segments are split at $\hat{\theta}_2^a = \frac{1}{3}$ and $\hat{\theta}_2^b = \frac{2}{3}$.

Compare this dynamic price discrimination game with the outcome under uniform pricing. Without the ability to condition second-period prices on first-period behavior, firms would offer the static prices in each period, $p_1^a = p_2^a = p_1^b = p_2^b = c + \tau$; consumers would pay a total of $(1 + \delta)(c + \tau)$ for the consumption stream, and no switching would arise. The absence of equilibrium switching under uniform pricing immediately

implies that price discrimination lowers social welfare, although the modeling assumption of inelastic demand again limits the generality of this conclusion.

Several additional results emerge from a simple comparison of uniform pricing and price discrimination. First, in the price discrimination game, the “loyal” consumers in the intervals $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$ do not switch and their present-value payment is the same as in the uniform setting: $p_1^j + \delta p_{2j}^j = c + \tau + \frac{\delta}{3}\tau + \delta(c + \frac{2}{3}\tau) = (1 + \delta)(c + \tau)$. Consumer surplus and profit for these intervals is unaffected by price discrimination. Second, the “poached” consumers in the interval $(\frac{1}{3}, \frac{2}{3})$ switch from one firm to the other. By revealed preference they could choose to be loyal but strictly prefer to switch firms; hence, consumer surplus increases for consumers with only moderate brand loyalties. Because such switching decreases social welfare, it follows that profits must decrease for these segments. The present-value price paid by these consumers for two periods of consumption is only $(1 + \delta)(c + \tau) - \frac{\delta}{3}\tau$, and hence lower than the price paid by loyal consumers. Price discrimination does not increase the present-value payment from any consumer segment and strictly reduces it to the middle segment, just as in models of all-out-competition.

The price path in the Fudenberg and Tirole (2000) model of product differentiation differs from the exogenous switching-cost model in Chen (1997b). In Fudenberg and Tirole (2000), prices *fall* over time as competition intensifies for the price-sensitive markets, while in Chen (1997b) prices *rise* over time as firms take advantage of buyer lock-in. One can understand the different dynamics by noting the different “network” effects present in each model.⁴⁶ In Chen (1997b), there is no network effect under price discrimination; second-period prices are independent of first-period market share. There is a network effect under uniform pricing, however, as a higher first-period market share generates a higher uniform price in the second-period. The reverse is the case in Fudenberg and Tirole (2000). Under price discrimination, there is a network effect as a larger first-period market share increases second-period prices. Under uniform pricing, there is no network effect as the second-period price is independent of first-period market share.

Villas-Boas (1999) studies a related but infinite-period, overlapping-generations model in which firms can only discriminate between returning customers and non-returning customers. Among these non-returning customers, a firm cannot distinguish between new consumers and customers who purchased from the rival firm in the first half of their economic lives. Whether this is a reasonable assumption depends upon the setting. If a firm wants to offer a lower price to new customers than to rival customers, this assumption may be plausible if masquerading as a new customer is possible. Of course, if a firm prefers to offer its rival customers the lower price, one might imagine a setting in which a consumer provides proof of purchase of a competing product; in that instance the assumption is less realistic and the model of Fudenberg and Tirole (2000) more appropriate. The steady-state results of Villas-Boas (1999) indicate

⁴⁶ See Farrell and Klemperer (2007) in this volume for a related discussion of networks.

that equilibrium prices are lower because each firm wants to attract the competitor's previous customers. Moreover, equilibrium prices decrease as consumers become more patient, due to increased indifference among marginal consumers over which product to buy initially. This movement toward indifference renders consumers more price sensitive, which in turn intensifies competition and makes customer retention more difficult. Finally, Villas-Boas (1999) demonstrates that, close to steady state, a greater previous-period market share results in a lower price to new customers and a higher price to existing customers. Thus, customer-recognition effects appear to make demand more inelastic in the previous period for reasons familiar to the other models of this section.

4.3. Purchase-history pricing with long-term commitment

Unlike the previous analyses which relied on short-term price agreements, we now ask what happens if a firm can write a long-term contract. The contract commits the firm to a second-period price to guarantee returning customers different terms than those offered to other customers. Banerjee and Summers (1987) and Caminal and Matutes (1990) were among the first to explore the use of long-term contracts to induce loyalty and generate endogenous switching costs.

Consider the setting of Caminal and Matutes (1990). As before, there are two firms and two periods of competition. The market in each period consists of a linear city of unit length, with firm a located at 0 and firm b located at 1, and consumers uniformly distributed across the interval. The difference with the previous models is that the location of each consumer is *independently distributed across periods*. Thus, a consumer's location in period 1, θ_1 , is statistically independent of the consumer's location in period 2, θ_2 . This independence assumption implies that there is no relevant second-period information contained in a consumer's first-period choice. It follows that if firms cannot commit to long-term prices, there is no value from price discrimination based upon purchase history.

Suppose, however, that price commitments are possible. The timing of the market game is as follows: First, firms simultaneously choose their first-period prices, p_1^j , and pre-commit (if they wish) to offer a second period price, p_{2j}^j , to customers purchasing in period 1 (i.e., period 1 customers are given an option contract for period 2). Consumers then decide from whom to purchase in period 1. At the start of the second period, each firm simultaneously chooses p_{2k}^j , $k \neq j$, which applies to non-returning consumers and returning customers if either $p_{2k}^j \leq p_{2j}^j$ or if firm j did not offer a price commitment. Caminal and Matutes (1990) demonstrate that subgame perfection requires that both firms commit to lower long-term prices for returning consumers. Calculating the subgame perfect equilibrium is straightforward, with the second-period poaching prices determined as functions of the second-period committed prices. These prices may be used to compute second-period profits and first-period market and derive equilibrium prices. Absent discounting, the equilibrium prices are $p_1^j = c + \frac{4\tau}{3}$, $p_{2j}^j = c - \frac{\tau}{3}$ and $p_{2k}^j = c + \frac{\tau}{3}$, $k \neq j$.

Caminal and Matutes (1990) find a few noteworthy results. First, equilibrium prices decline over time. Remarkably, the second-period commitment price is below even marginal cost. The reasoning of this is subtle. Suppose, for example, that firm b 's poaching price in the second period was independent of firm a 's second-period loyalty price. Because the consumer's second-period location is unknown at the time they enter into a long-term contract, firm a maximizes the joint surplus of a consumer and itself by setting $p_{2j}^j = c$ and pricing efficiently in the second period. Given that firm b 's poaching price in reality *does* depend positively on firm a 's second-period loyalty price, firm a can obtain a first-order gain in joint surplus by reducing its price slightly below cost and thereby reducing firm b 's second-period price. This slight price reduction incurs only a corresponding second-order loss in surplus since pricing was originally at the efficient level. Hence, a firm can commit to a follow-on price below marginal cost in the second-period as a way to increase the expected surplus going to the consumer, and thereby raise the attractiveness of purchasing from the firm initially. Price commitment in this sense has a similar flavor to the models of Diamond and Maskin (1979) and Aghion and Bolton (1987), in which contractual commitments are used to extract a better price from a third party. Of course, as Caminal and Matutes (1990) confirm, when both firms undertake this strategy simultaneously, profits fall relative to the no-commitment case and firms are worse off. Welfare is also lower as too little switching takes place from a social viewpoint because lock-in is always socially inefficient.

What are the effects of the presence of this commitment strategy? As is by now a familiar theme, although an individual firm will benefit from committing to a declining price path for returning customers, the firms are collectively worse off with the ability to write long-term price contracts. With commitment, there is also too much lock-in or inertia in the second-period allocations. Without commitment, social welfare would therefore be higher, as consumers would allocate themselves to firms over time to minimize transportation costs. The endogenous switching costs (created by the declining price path for returning consumers) decrease social welfare. As before, because market demand is inelastic in this model, there is an inherent bias against price discrimination, so we must carefully interpret the welfare conclusions.⁴⁷

⁴⁷ Caminal and Matutes (1990) also consider a distinct game in which firms' strategies allow only for commitments to discounts (or coupons) rather than to particular prices. In this setting, a similar decreasing price path emerges for continuing consumers, and profits for firms are higher than with commitment to prices. Indeed, the profits are higher than if no commitments are allowed. The difference arises because committed prices do not have an impact on second-period profits from non-returning customers; this is not the case with committed discounts. With committed discounts, firms are reluctant to cut prices to non-returning customers, so second-period competition is less intense. This finding relates to that of Banerjee and Summers (1987), who show in a homogeneous product market that commitments to discounts for returning customers can be a collusive device which raises second-period prices to the monopoly level. Because profits are higher in the discount game, we might expect firms to choose such strategies in a meta-game that offers a choice between committed discounts and committed prices. Caminal and Matutes (1990) analyze this meta-game and conclude that, unfortunately for the firms, the price-commitment game is the equilibrium outcome. The analysis of Caminal and Matutes (1990) demonstrates that price discrimination over purchase history can generate

Closely related to Caminal and Matutes (1990), Fudenberg and Tirole (2000) also consider an environment in which price commitments can be made through long-term contracts, using an interesting variation in which consumer preferences are *fixed* across periods (i.e., location θ between firms a and b does not change). In this setting, there are no exogenous switching costs, but long-term contracts with breach penalties are offered which introduce endogenous switching costs in the sense of Caminal and Matutes (1990). While such contracts could effectively lock consumers into a firm and prevent poaching, the firms choose to offer both long-term and spot contracts in equilibrium so as to segment the marketplace. In equilibrium, consumers with strong preferences for one firm purchase long-term contracts, while those with weaker preferences select short-term contracts and switch suppliers in the second period. The firm utilizes long-term contracts to generate lower poaching prices, which benefit consumers located near the center of the market. The firm can extract concessions in the first period by locking in some customers with long-term contracts and thereby generate more aggressive second-period pricing [similar in spirit to Caminal and Matutes (1990)]. The long-term contracts generate a single-crossing property that segments first-period consumers: in equilibrium, consumers located near the middle of the market are more willing to purchase short-term contracts and switch in the second period to a lower-priced rival than consumers located at the extreme.

It is worth noting that a multi-plant monopolist would accomplish a similar sorting by selling long-term contracts for aa , bb and the switching bundles ab and ba . The monopolist can therefore segment the market, charging higher prices to the non-switchers and lower prices to the consumers who are willing to switch. The optimal amount of monopoly switching in the second period (given a uniform distribution) can be shown to be one-half of the consumers. Here, long-term contracts serve a similar segmentation function. Interestingly, Fudenberg and Tirole (2000) show that if consumers are uniformly distributed, one-fourth of the consumers will switch in the long-term contracting equilibrium with duopoly – less switching than in the case of short-term contracts alone, and still less switching than under monopoly. Hence, given price discrimination is allowed, allowing firms to use long-term contracts improves social welfare. This finding emerges because demand does not change across periods, in contrast to Caminal and Matutes (1990), who find that when preferences vary over time, long-term price commitments lead to too much lock-in and hence reduce social welfare.

5. Intrapersonal price discrimination

Intrapersonal price discrimination has received very little attention, partly because in the context of monopoly, the models and results are economically immediate. For example,

endogenous switching costs with declining price paths (loyalty rewards) and too much lock-in, assuming that demand information in the first period is independent of the second period.

consider a consumer with a known demand curve, $p = D(q)$, and a monopolist with constant marginal (and average) cost of production equal to c . Let q^* be the unique solution to $c = D(q^*)$. The monopolist increases profits by offering any of a host of equally optimal but distinct mechanisms that encourage efficient consumption on the margin: a fixed bundle of q^* units at a bundle price of $\int_0^{q^*} D(z) dz$; a fully non-linear tariff of $P(q) = D(q)$; or a two-part tariff equal to $P(q) = v(c) + cq$, where $v(p) \equiv \max_q u(q) - pq$ is the consumer's indirect utility of consuming at the linear price p . In all examples, the monopolist effectively price discriminates in an intrapersonal manner: Different prices are charged to the same consumer for different units because the consumer's marginal values for those units vary according to consumption. These outcomes closely relate to third-degree price discrimination because the different units of consumption can be thought of as distinguishable market segments. Because consumers are homogeneous within each market – the reservation value is $D(q)$ in the q th unit market – price discrimination is perfect and social welfare is maximized. Note that if our downward sloping demand curve were instead generated by a continuous distribution of heterogeneous consumers with unit demands, then perfect price discrimination is no longer possible. There is no effective way to capture the infra-marginal surplus in this interpersonal analogue.

When markets are imperfectly competitive, intrapersonal price discrimination using non-linear prices allows firms to provide more efficiently a given level of consumer surplus while covering any fixed, per-consumer, production costs. This efficiency suggests that intrapersonal price discrimination may raise social welfare when firms compete. Following [Armstrong and Vickers \(2001\)](#), we address this possibility and related issues of consumer surplus and industry profit in the simplest discrete-choice setting in which there is a single market segment with homogeneous consumers.⁴⁸

Suppose that there are two firms, $j = a, b$, and each is allowed to offer price schedules, $P_j(q_j)$, chosen from the set \mathcal{P} . Any restrictions on pricing, such as a requirement of uniform, per-unit prices, are embedded in \mathcal{P} . We assume that consumers make all of their purchases from a single firm. This one-stop-shopping assumption implies that a consumer evaluates his utility from each firm and chooses either the firm which generates the greatest utility or not to purchase at all.⁴⁹ A consumer who buys from firm j obtains indirect utility of

$$v_j \equiv \max_q u(q) - P_j(q).$$

Define $\pi(v)$ to be the maximal expected per-consumer profit that firm j can make, while choosing $P_j \in \mathcal{P}$ and generating an indirect utility of v for each participating

⁴⁸ [Armstrong and Vickers \(2001\)](#) also study the important settings of interpersonal third-degree price discrimination (discussed in Section 3.3) and the case of unobservable heterogeneity considered in Section 6.

⁴⁹ [Holton \(1957\)](#) provides an early discussion of one-stop shopping between supermarkets practicing price discrimination. See also [Bliss \(1988\)](#) for a model of one-stop shopping.

consumer:

$$\pi(v) \equiv \max_{P_j \in \mathcal{P}} P_j(q) - C(q), \quad \text{such that } \max_q u(q) - P_j(q) = v.$$

Note that as fewer restrictions are placed on \mathcal{P} and the set increases in size, $\pi(v)$ weakly increases. Thus, the per-consumer profit with unfettered price discrimination, $\pi^{\text{pd}}(v)$, will typically exceed the per-consumer profit when prices are restricted to be uniform across units, $\pi^u(v)$.

Following the discrete-choice literature,⁵⁰ we can model duopoly product differentiation by assuming that each consumer's net utility from choosing firm j is the consumer's indirect utility plus an additive, firm-specific, fixed effect, $v_j + \varepsilon_j$; the outside option of no purchase is normalized to 0. For any joint distribution of the additive disturbance terms across firms, there exists a market share function, $\phi^a(v_a, v_b)$, which gives the probability that a given consumer purchases from firm a as a function of the indirect utilities offered by the two firms. Let $\phi^b(v_a, v_b) \equiv \phi^a(v_b, v_a)$ give the corresponding symmetric probability of purchase from firm b , where $\phi^a(v_a, v_b) + \phi^b(v_a, v_b) \leq 1$. With this notation in hand, we can model firms competing in utility space rather than prices. Firm a maximizes $\phi^a(v_a, v_b)\pi(v_a)$, taking v_b as given, and similarly for firm b . Armstrong and Vickers (2001) show with a few regularity assumptions that a symmetric equilibrium is given by each firm choosing a v to maximize $\sigma(v) + \log \pi(v)$, where $\sigma(v)$ is entirely determined by the structure of the market share functions.

Remarkably, one can separate the marginal effects of v on market share from its effect on per-consumer profit, which provides a powerful tool to understand the impact of intrapersonal price discrimination in competitive settings. For example, to model the competition for market share, $\phi(v_a, v_b)$, assume that the duopolists are situated at either end of a linear Hotelling-style market with transportation costs, τ , and uniformly distributed consumers.⁵¹ Take the simplest case where per-consumer costs are $C(q) = cq + k$. We then consider two cases: when $k > 0$, there is a per-consumer cost of service; and when $k = 0$, there are constant returns to scale in serving a consumer. We define \bar{v} to be the highest level of indirect utility that can be given to a consumer while earning nonnegative profits; formally,

$$\pi(\bar{v}) = 0 \quad \text{and} \quad \pi(v) < 0 \quad \text{for all } v > \bar{v}.$$

When there is a fixed-cost per consumer, $k > 0$, then $\bar{v} = v(c) - k$ when two-part tariffs are allowed, but $\bar{v} < v(c) - k$ when prices must be uniform. Thus, when $k > 0$ and per-consumer fixed costs exist, it follows that price discrimination generates greater

⁵⁰ Anderson, de Palma and Thisse (1992) provide a thorough survey of the discrete-choice literature.

⁵¹ Formally, this framework violates a technical assumption used in Armstrong and Vickers's (2001) separation theorem, due to the kinked demand curve inherent in the Hotelling model. It is still true, however, that the equilibrium utility maximizes $\sigma(v) + \log \pi(v)$ if the market is covered. Here, $\sigma(v) = v/\tau$. Armstrong and Vickers (2001) also demonstrate that the Hotelling framework approximates the discrete-choice Logit framework when competition is strong.

utility than does uniform pricing: $\bar{v}^{pd} > \bar{v}^u$. In such a setting, Armstrong and Vickers (2001) prove that allowing price discrimination increases consumer surplus and welfare relative to uniform pricing. As competition intensifies (i.e., $\tau \rightarrow 0$), firms attract consumers only by offering them utility close to the maximal zero-profit level. Because the relative loss in profits from a gain in indirect utility is never more than one-to-one, social surplus also increases. In a related model with free entry and firms competing equidistant on a Salop-style circular market, Armstrong and Vickers (2001) similarly demonstrate that price discrimination increases consumer surplus (which equals welfare) relative to uniform pricing. Because welfare increases under price discrimination, it follows that output must increase.

In the case when $k = 0$ and there is no per-consumer fixed cost, it follows that $\bar{v}^{pd} = \bar{v}^u = v(c)$. Using a more subtle economic argument that requires Taylor expansions around $\tau = 0$, Armstrong and Vickers (2001) find that as competition intensifies (i.e., $\tau \rightarrow 0$), price discrimination increases welfare and profits, but this time at the expense of consumer surplus. In the free-entry analog on a circular market, they demonstrate that price discrimination again increases welfare and consumer surplus (profits are zero), relative to uniform pricing. It is more difficult to capture the economic intuition for these results, given that the analysis relies on second-order terms and $\tau \approx 0$. Because $\pi^{pd}(\bar{v}) = \pi^u(\bar{v})$ and $\pi^{pd'}(\bar{v}) = \pi^{u'}(\bar{v})$, it can be shown that the second-order terms $\pi^{pd''}(\bar{v}) > \pi^{u''}(\bar{v})$ in the Taylor expansions drive the result. Taken together with the case for $k > 0$, the findings suggest that intrapersonal price discrimination is welfare-enhancing when competition is strong.

6. Non-linear pricing (second-degree price discrimination)

Unlike the setting of third-degree price discrimination, indirect (second-degree) discrimination relies on self-selection constraints, thus introducing an entirely new set of competitive issues.

In what follows, firms compete via price schedules of the form $P_j(q_j)$, and consumers choose which (if any) firms to patronize and which product(s) from the offered lines they will purchase. To model imperfect competition, we assume the product lines are differentiated.⁵² The theoretical literature on second-degree price discrimination under imperfect competition has largely focused on characterizing equilibrium schedules and

⁵² Other papers have modeled imperfect competition and non-linear pricing for homogeneous products by restricting firms to the choice of quantities (or market shares) as strategic variables, but we do not consider these approaches in this survey. Gal-Or (1983), De Fraja (1996), and Johnson and Myatt (2003) all consider the setting in which firms choose quantities of each quality level, and the market price schedule is set by a Walrasian auctioneer so as to sell the entire quantity of each quality. In a related spirit, Oren, Smith and Wilson (1983) consider two distinct models of imperfect competition with homogeneous goods. In the first, each firm commits to the market share it serves for each quality level; in the second model, each firm commits to the market share it serves for each consumer type.

the consequences of competition on efficiency; less attention has been spent on the desirability of enforcing uniform pricing in these environments. This is due in part to the extra technical complexity of second-degree price discrimination, and, to a lesser degree, to the impracticality of requiring uniform pricing when q refers to quality rather than quantity.

The variety of consumer preferences and competitive environments make it useful to distinguish a few cases. Two possible equilibrium configurations can arise: a consumer may purchase exclusively from one firm (referred to in the contract theory literature as *exclusive agency*) or may purchase different products from multiple firms (referred to as *common agency*). Second, within each agency setting, there are several possibilities regarding the unobservable heterogeneity of consumers. The two most common forms are what we will call *vertical* and *horizontal* heterogeneity. In the former, the consumer's marginal preferences for q (and absolute preferences for participating) are increasing in θ for each firm; in the latter, the consumer's marginal preferences for q and absolute preferences for participating are monotonic in θ , but the direction varies across firms with the result that a high-demand type for firm j is a low-demand type for firm k and conversely.⁵³ Under these definitions, vertical heterogeneity implies that firms agree in their ranking of type from high demand to low demand; under horizontal heterogeneity, two firms have reversed ranking for consumer types. In this sense, the taxonomy is similar in spirit to best-response symmetry and asymmetry under third-degree price discrimination.

Ideally, a model of competition among second-degree price-discriminating firms should incorporate both vertical and horizontal dimensions of heterogeneity to capture both a common ranking of marginal valuations for quality among consumers (holding brand preferences fixed) and a variety of brand preferences (holding quality valuations fixed). Unfortunately, multidimensional self-selection models are considerably more difficult to study, as they introduce additional economic and technical subtleties.⁵⁴ As a result, the literature has either relied upon one-dimensional models (either vertical or horizontal) for precise analytic results [e.g., Spulber (1989), Martimort (1992, 1996), Stole (1991), Stole (1995), Martimort and Stole (2005)], numerical simulations of multidimensional models [e.g., Borenstein (1985)], or further restrictions on preferences to simplify multidimensional settings to the point of analytical tractability [e.g., Armstrong and Vickers (2001), Rochet and Stole (2002), Ivaldi and Martimort (1994)]. While there are few general results across these different models, the range of approaches

⁵³ Formally, if preferences for firm j 's goods are represented by $u^j(q_j, \theta)$, then vertical heterogeneity exists when $u_\theta^j > 0$ and $u_{q\theta}^j > 0$ for each j . Horizontal heterogeneity is said to exist between firms j and k if $u_{q\theta}^j > 0 > u_{q\theta}^k$ and $u_\theta^j > 0 > u_\theta^k$. Note that this notion of vertical heterogeneity should not be confused with the pure vertical differentiation preferences described in Gabszewicz and Thisse (1979, 1980) and Shaked and Sutton (1982, 1983), which require that all potential qualities are ranked in the same way by every consumer type when products are priced at cost.

⁵⁴ See Armstrong and Rochet (1999) and Rochet and Stole (2003) for surveys of multidimensional screening models.

successfully underscores a few recurring economic themes which we will illustrate in this section.

Before we survey the various approaches taken in the literature, we first review monopoly second-degree price discrimination to provide a benchmark and to introduce notation.

6.1. Benchmark: monopoly second-degree price discrimination

In the simplest monopoly model, the consumer has preferences $u(q, \theta) - P(q)$ over combinations of q and price, $P(q)$, where the consumer's one-dimensional type, θ , is distributed on some interval $\Theta = [\theta_0, \theta_1]$ according to the distribution and density functions, $F(\theta)$ and $f(\theta)$, respectively. Assume that the outside option of not consuming is zero and that the firm faces a per-consumer, convex cost of quality equal to $C(q)$. As in Mussa and Rosen (1978), we take q to represent quality, but it could equally well represent quantities.⁵⁵ We also assume that the consumer's preferences satisfy the standard single-crossing property that $u_{q\theta}(q, \theta) > 0$ and utility is increasing in type, $u_\theta(q, \theta) > 0$. In terms of consumer demand curves for q indexed by type, $p = D(q, \theta)$, single-crossing is equivalent to assuming the demand curves are nested in θ , with higher types exhibiting a greater willingness to pay for every increment of q .

The firm chooses the price schedule, $P(q)$, to maximize expected profits, given that consumers will select from the schedule to maximize individual utilities. Solving for the optimal price schedule is straightforward. For any $P(q)$, the firm can characterize the associated optimal purchase and surplus for a consumer of type θ as

$$q(\theta) \equiv \arg \max_q u(q, \theta) - P(q),$$

$$v(\theta) \equiv \max_q u(q, \theta) - P(q).$$

Expected profits can be written in terms of expected revenue less cost, or in terms of expected total surplus less consumer surplus:

$$\int_{\theta_0}^{\theta_1} (P(q(\theta)) - C(q(\theta))) dF(\theta) = \int_{\theta_0}^{\theta_1} (u(q(\theta), \theta) - C(q(\theta)) - v(\theta)) dF(\theta).$$

The monopolist cannot arbitrarily choose $q(\theta)$ and $v(\theta)$, however, as the consumer's ability to choose q must be respected. Following standard arguments in the self-selection literature, we know that such incentive compatibility requires that $q(\theta)$ weakly increases in θ and that the consumer's marginal surplus is $v'(\theta) = u_\theta(q(\theta), \theta) > 0$. The requirement that the consumer wishes to participate, $v(\theta) \geq 0$, will be satisfied if and only if $v(\theta_0) \geq 0$, given that $v'(\theta) > 0$. Fortunately, for any surplus and quality functions that satisfy the incentive-compatibility and participation conditions, a unique price schedule (up to a constant) exists that implements $q(\theta)$.

⁵⁵ Maskin and Riley (1984) consider a more general model with non-linear pricing over quantities to explore, among other things, whether quantity discounts are optimal.

Integrating by parts and substituting for $v'(\theta)$ converts the firm's constrained-maximization program over $\{q(\theta), v(\theta)\}$ to a simpler program over $q(\theta)$ and $v(\theta_0)$. In short, the firm chooses $q(\theta)$ to maximize

$$\int_{\theta_0}^{\theta_1} \left(u(q(\theta), \theta) - C(q(\theta)) - \frac{1 - F(\theta)}{f(\theta)} u_{\theta}(q(\theta), \theta) - v(\theta_0) \right) dF(\theta),$$

subject to $q(\theta)$ nondecreasing and $v(\theta_0) \geq 0$.

Assuming that the firm finds it profitable to serve the entire distribution of consumers, it will choose $v(\theta_0) = 0$ – a corner solution to the optimization program.⁵⁶ The integrand of firm's objective function is defined by

$$\Lambda(q, \theta) \equiv u(q, \theta) - C(q) - \frac{1 - F(\theta)}{f(\theta)} u_{\theta}(q, \theta).$$

This “virtual” profit function gives the total surplus less the consumer's information rents for a fixed type, θ . Choosing $q(\theta)$ to maximize $\Lambda(q, \theta)$ pointwise over θ will simultaneously maximize its expected value over θ . If this function is strictly quasiconcave (a reasonable assumption in most contexts), then the optimal $q(\theta)$ is determined by $\Lambda_q(q(\theta), \theta) = 0$ for every θ . If $\Lambda(q, \theta)$ is also supermodular (i.e., $\Lambda_{q\theta}(q, \theta) \geq 0$) – an assumption that is satisfied for a wide variety of distributions and preferences – then the resulting function $q(\theta)$ is weakly increasing in θ .⁵⁷ Hence, with a few regularity conditions, we have the firm's optimal choice. After setting $v(\theta_0) = 0$ and determining $q(\theta)$, constructing the unique price schedule is straightforward: one recovers $v(\theta) = v(\theta_0) + \int_{\theta_0}^{\theta} u_{\theta}(q(\theta), \theta) d\theta$ and then constructs $P(q)$ from the equation $u(q(\theta), \theta) - P(q(\theta)) = v(\theta)$.

Because we are primarily interested in the consumption distortions introduced by the monopolist, we are more interested in $q(\theta)$ than $P(q)$. To understand the distortion, consider the first-order condition, $\Lambda_q(q(\theta), \theta) = 0$:

$$u_q(q(\theta), \theta) - C_q(q(\theta)) = \frac{1 - F(\theta)}{f(\theta)} u_{q\theta}(q(\theta), \theta) \geq 0. \quad (1)$$

In words, the marginal social benefit of increasing q (the marginal increase in the size of the pie) is set equal to the marginal loss from increased consumer surplus and reduced infra-marginal profits (a smaller slice of the pie). Alternatively, in the form $f(\theta)(u_q - C_q) = (1 - F(\theta))u_{q\theta}$, we see that the gain from increasing quality for some type θ is the probability, $f(\theta)$, of that type arising, multiplied by the increase in surplus which the

⁵⁶ For sufficiently large heterogeneity, it is possible that the firm will wish to serve a proper subset of types. Here, the lowest type served, θ_0^* , is determined by the point where the virtual profit (defined below) of the monopolist goes from positive to negative: $\Lambda(q(\theta_0^*), \theta_0^*) = 0$.

⁵⁷ For example, if preferences are quadratic and $(1 - F(\theta))/f(\theta)$ is non-increasing in θ , then $\Lambda(q, \theta)$ is supermodular. When $\Lambda(q, \theta)$ is not supermodular, one must employ control-theoretic techniques to maximize, subject to the monotonicity constraint. This “ironing” procedure is explained in Fudenberg and Tirole (1991, ch. 7).

marginal quality generates, $(u_q - C_q)$. The loss arises from all higher type consumers, $1 - F(\theta)$, who obtain more surplus by the amount $u_{q\theta}$. Thus, we have the marginal-versus-inframarginal tradeoff that is familiar to the classic monopolist: the marginal profit from selling to one more consumer must be set against the lowered price given to the higher-demand, inframarginal customers.

In standard models of price- or quantity-setting oligopolists, competition reduces the significance of the inframarginal term and fewer distortions arise. One might conjecture that the presence of competition should have the same general effect in markets with non-linear pricing – reducing the impact of the infra-marginal effect (and thereby reducing the distortions from market power). As we will see below, this is the case for a large class of models.

Two final remarks on monopoly price discrimination are helpful. First, the above model was one of vertical differentiation; consumers' willingness to pay increases in their marginal valuation of quality, θ , and the firm makes more profit per customer on the high types than on the low types. One could instead consider a model of horizontal differentiation with little difference in the character of the distortions. For example, suppose that consumer types are distributed over the positive real numbers, and a type represents the distance of the consumer to the monopolist firm. Here, closer consumers (low θ 's) take on the role of valued, high-demand customers, so $u_\theta < 0$ and $u_{q\theta} < 0$. The analysis above executes with very minor modifications. Now, all consumers but the closest to the firm will have downward distortions in their quality allocation, and the first-order condition will be given by an analogue of the vertical uncertainty program:

$$u_q(q(\theta), \theta) - C_q(q(\theta)) = -\frac{F(\theta)}{f(\theta)} u_{q\theta}(q(\theta), \theta) \geq 0.$$

Hence, there is nothing conceptually distinct about horizontal or vertical preference heterogeneity in the context of monopoly, providing the relevant single-crossing property is satisfied.

Second, it is worth noting that a consumer's relevant "type" is completely summarized by the consumer's demand curve; therefore, we should be able to find similar conclusions looking only at a distribution of demand functions, indexed by type. Consider again the case of vertical heterogeneity, and denote $p = D(q, \theta)$ as a type- θ consumer's demand curve for quality. By definition, $D(q, \theta) \equiv u_q(q, \theta)$. The relevant condition for profit maximization is now

$$D(q(\theta), \theta) - C_q(q(\theta)) = \frac{1 - F(\theta)}{f(\theta)} D_\theta(q(\theta), \theta).$$

Rearranging, this can be written in the more familiar form

$$\frac{P'(q(\theta)) - C_q(q(\theta))}{P'(q(\theta))} = \frac{1}{\eta(q(\theta), \theta)},$$

where $\eta = \frac{1-F}{f} \frac{D_\theta}{D}$ is the relevant elasticity of marginal demand for the $q(\theta)$ marginal unit. In the elasticity form, the intuitive connection between non-linear pricing

and classic monopoly pricing is clear. In a large variety of competitive non-linear pricing models, the effect of competition is to increase this elasticity and hence reduce marginal distortions. To understand how this elasticity formula changes under competition, we separately examine the settings in which a consumer makes all purchases from a single firm, or commonly purchases from multiple firms.

6.2. Non-linear pricing with one-stop shopping

Suppose that in equilibrium, consumers purchase from at most one firm. For example, each consumer may desire at most one automobile, but may desire a variety of quality-improving extras (air conditioning, high performance stereo, luxury trim, etc.) which must be supplied by the same seller. In this setting of one-shop shopping, firms compete for each consumer's patronage. One can think of the consumer's decision in two stages: first, the consumer assigns an indirect utility of purchasing from each firm's product's line and associated price schedule, and second, the consumer visits the firm with the highest indirect utility (providing it is nonnegative) and makes his purchase accordingly.

6.2.1. One-dimensional models of heterogeneity

Assume for the present that all uncertainty in the above setting is contained in a one-dimensional parameter, θ , and that preferences for each firm j 's products are given by $u^j(q_j, \theta) - P_j(q_j)$, $j = 1, \dots, n$. Given the offered schedules, the indirect utility of purchasing from each firm j is

$$v_j(\theta) = \max_{q_j} u^j(q_j, \theta) - P_j(q_j).$$

We let $v_0(\theta) = 0$ represent utility associated with not purchasing from any firm. Calculating the indirect utilities, firm j can derive the best alternative for each consumer of type θ , relative to the firm's offer:

$$\underline{v}_j(\theta) \equiv \max_{k \neq j} v_k(\theta),$$

where the utility from no purchase is included in the maximand on the right. The competitive environment, from firm j 's point of view, is entirely contained in the description of $\underline{v}_j(\theta)$. The best response of firm j is the same as a monopolist which faces a consumer with utility $u(q, \theta) - P(q)$ and an outside option of $\underline{v}_j(\theta)$.

This monopoly restatement of firm i 's problem makes clear a new difficulty: the outside option depends on θ . Economically, the presence of θ in the outside option means that it is no longer clear which consumer types will be marginally attracted to a firm's rival.⁵⁸ Fortunately, some guidance is provided as there exists a connection between

⁵⁸ A number of theoretical contributions in the incentives literature have developed the methodology of type-dependent participation constraints; see, for example, Lewis and Sappington (1989), Maggi and Rodriguez-Clare (1995) and Jullien (2000).

the nature of the participation constraint and the form of preference heterogeneity – horizontal or vertical.⁵⁹

• *Horizontal heterogeneity* Consider a setting in which consumers are located between two firms such that closer consumers have not only lower transportation costs, but also a higher marginal utility of quality. For example, consumers without strong preferences between an Apple- or Windows-based desktop computer are also not willing to pay as much for faster processors or extra software packages in their preferred product line; the reverse is true for someone with strong brand preferences. This is a setting of horizontal heterogeneity. Formally, let θ represent a consumer's distance to firm 1 (and her "closeness" to firm 2); we assume that $u_{\theta}^1(q, \theta) < 0 < u_{\theta}^2(q, \theta)$. Because greater distance lowers the marginal utility of quality, we also have $u_{q\theta}^1(q, \theta) < 0 < u_{q\theta}^2(q, \theta)$. Holding q fixed, a consumer that is close to firm 1 and far from firm 2 obtains higher utility from firm 1 than firm 2, and has a higher marginal valuation of firm 1's quality than firm 2's. It follows that a low- θ (respectively, high- θ) consumer has a higher demand for firm 1's (respectively, firm 2's) product line.

An early and simple model of non-linear pricing by oligopolists in a setting of horizontal heterogeneity is developed in Spulber (1989), which we follow here.⁶⁰ Firms are evenly spaced on a circular market of unit size, and consumers' types are simply their locations on the market. Consider a consumer located between two firms, with the left firm located at 0 and the right one located at $\frac{1}{n}$. The consumer's utility from the left firm at price p is taken to be $u^1 = (z - \theta)q - P_1(q)$ and the utility derived from the right firm is $u^2 = (z - (\frac{1}{n} - \theta))q - P_2(q)$. Here, the base value of consumption, z , is known, but the consumer's location in "brand" space is private information. There is a single-crossing property in (q, θ) for each firm: nearer consumers enjoy a higher margin from consuming q .

Consider firm 2's problem. Taking $P_1(q)$ as fixed, we have $\underline{v}_2(\theta) = \max\{0, v_1(\theta)\}$, where $v_1(\theta) = \max_q(z - \theta)q - P_1(q)$. Using the envelope theorem, we know that $v_1'(\theta) = -q_1(\theta) \leq 0$, so $\underline{v}_2(\theta)$ is decreasing. As a result, if the marginal type, $\tilde{\theta}$, is indifferent between the two firms, then firm 1 obtains market share $[0, \tilde{\theta})$ and firm 2 obtains the share $(\tilde{\theta}, \frac{1}{n}]$. This partition implies that the determination of market share and the allocation of quality are separable problems. Price competition between the local duopolists is *entirely* focused on the marginal consumer. Here, each firm will trade off the gain derived from the additional market share captured by raising this marginal consumer's utility against the cost of lowering prices to all inframarginal consumers.

⁵⁹ Several papers have drawn upon this distinction in one form or another; see, for example, Borenstein (1985), Katz (1985), and Stole (1995).

⁶⁰ Spulber (1984) also considers spatial non-linear pricing with imperfect competition, but models imperfect competition by assuming firms are local monopolists within a given market radius and that the market radius is determined by a zero-profit condition. Unlike Spulber (1989), this approach to spatial competition ignores the interesting price effects of competition on the boundaries. Norman (1981) considers third-degree price discrimination with competition under a similar model of spatial competition.

This classic tradeoff determines the value of $\underline{v} = v_1(\tilde{\theta}) = v_2(\tilde{\theta})$. Given the equilibrium partition and utility of the marginal consumer, \underline{v} , each firm will allocate $q(\theta)$ as if they were a second-degree price-discriminating monopolist constrained to offer $v(\theta) \geq \underline{v}$. As Spulber (1989) emphasizes, the resulting quality allocations are equivalent to those of a monopolist operating over the given market shares. Given the quality allocation is unchanged between competition and monopoly, the marginal price schedules are the same whether the market structure is an n -firm oligopoly or an n -plant monopolist. Only the *level* of prices will be lower under competition, leaving positive surplus to the marginal “worst-type” consumer.

A few remarks are in order. First, a multi-product monopolist may choose not to cover the market, while every consumer would otherwise be served under oligopoly, so we must be careful to consider market coverage in making welfare conclusions. Second, fixing the number of distinct product lines, n , and assuming that the market is covered under monopoly and oligopoly, social welfare is unaffected by the market structure. Under competition, consumer surplus is higher and profits are lower, but the aggregate surplus is unchanged from a multi-product monopoly selling the same n product lines. Third, as the number of product lines, n , increases and more brands become available, the distance between brands decreases. The marginal customer then moves closer to her nearest brand, leading to a reduction in quality distortions. Of course, more product lines typically come at a social cost, which raises a final point: In a free-entry model with fixed per-plant costs of production. Social welfare may decrease from excessive entry.

- *Vertical heterogeneity* Suppose instead that unobservable heterogeneity is vertical rather than horizontal; i.e., every firm ranks the consumer types identically. Formally, $u_{q\theta}^j(q_j, \theta) > 0$ and $u_{\theta}^j(q_j, \theta) > 0$ for $j = 1, \dots, n$. Given that all firms rank the consumer types equivalently and θ does not represent differentiated tastes for brands, the source of product differentiation is not immediate. Two approaches to modeling imperfect competition emerge in such a setting.⁶¹ The first assumes firms have different comparative advantages for serving customer segments; e.g., Stole (1995). The second assumes that the firms are ex ante symmetric in their productive capabilities, but in an initial stage the firms commit to a range of qualities before the choosing prices, generating endogenous comparative advantages that soften price competition; e.g., Champsaur and Rochet (1989). We consider each approach in turn.

Consider a duopoly where firm 1 has a commonly known comparative advantage over firm 2 for an identifiable consumer segment. As a simple example, suppose that a consumer obtains $u^1(q_1, \theta) = \theta q_1 + z - P_1(q_1)$ from firm 1 and $u^2(q_2, \theta) = \theta q_2 - P_2(q_2)$

⁶¹ Locay and Rodriguez (1992) take a third approach. They assume that buyers must consume goods in groups (e.g., a family or club), and that the group chooses a firm (one-stop shopping) based on some group-welfare norm. Competition occurs between firms for groups and drives profits to zero. With zero profits there remains intra-group price discrimination to allocate costs.

from firm 2, although the cost of production is $C(q) = \frac{1}{2}q^2$ for both firms. In equilibrium, firm 2 offers $P_2(q) = C(q)$ and the consumer's associated indirect utility becomes $\underline{v}_1(\theta) = \frac{1}{2}\theta^2$. Firm 1, in order to make any sales, must offer at least $\frac{1}{2}\theta^2$ to any type it wishes to serve. Straightforward techniques of optimization with type-dependent participation constraints verify that in this equilibrium, $P_1(q) = C(q) + z$, providing that the consumer who is indifferent between the two firms chooses to purchase from firm 1. In contrast to the setting of horizontal heterogeneity, the participation constraint binds for every type of consumer. One may also note that the equilibrium allocation is efficient; it is known with certainty that firm 1 extracts all the consumer's residual surplus, z . Consumers nonetheless obtain considerable information rents given the strong competitive pressures from firm 2.

This model of vertical heterogeneity is perhaps too simple, because neither firm has a comparative advantage on the margin of quality, and thus the firms are perfectly competitive. Firm 1 only extracts rents on its comparative advantage, the additive component of z , which can be extracted without distortion given full information about z . A more general model would allow variations in the marginal comparative advantages of the firms. Along these lines of inquiry, Stole (1995) assumes that consumers are located on a unit circle with n evenly-spaced oligopolists *but* each consumer's location is observable so that firms can offer delivered, non-linear prices conditioned on location. For a consumer segment located at $x \in [0, \frac{1}{n}]$, suppose that the utility of consuming from the first firm is $u^1 = (\theta - x)q - P_1(q)$, while the utility of consuming from the second firm is $u^2 = (\theta - (\frac{1}{n} - x))q - P_2(q)$. Firms furthermore have symmetric costs of production, say $C(q) = \frac{1}{2}q^2$. A reasonable conjecture would have the consumer purchasing from the nearest firm (say, firm 1), while the more distant firm offers its product line at cost and makes no sales, $P_2(q) = C(q)$. Now it is less clear for which types the participation constraint will bind. The closer firm now faces a competitive pressure which guarantees consumer surplus of at least $\underline{v}_1(\theta) = \frac{1}{2}(\theta - (\frac{1}{n} - x))^2$, which is increasing in θ .

For the moment, suppose the nearer firm offers the monopoly quality schedule, assuming that the outside option only matters for the lowest type, θ_0 , and that θ is distributed uniformly on $[\theta_0, \theta_1]$. Straightforward calculations reveal that $q_1(\theta) = 2\theta - \theta_1 - x$. The envelope theorem implies that the slope of the consumer's indirect utility function from firm 1 is $v'_1(\theta) = q_1(\theta)$; consequently, there must be parameter values (e.g., $\frac{1}{n} > 2x + \theta_1 - \theta_0$) such that the participation constraint binds for only the lowest type. More generally, an interior value of $\hat{\theta}$ exists such that the participation constraint binds for all $\theta < \hat{\theta}$, and is slack otherwise. When this happens, it must be the case that $q_1(\theta) = \underline{v}'_1(\theta)$ over the determined interval, which in turn requires that q_1 is less distorted than it would be without the binding constraint. While the calculation of this interval relies upon control-theoretic techniques, it helps us understand how competition affects distortions. Here, the rival firm's offer has an additional competitive effect, actually increasing quality for the lower end of consumer types. When the number of firms increases, the firms become closer and the rival offer becomes more attractive, binding over a larger interval as the differentiation between firms decreases.

While these two simple examples of vertical heterogeneity may be of applied interest, they fail to adequately portray the richness of results that can arise in a setting of vertical heterogeneity when firms have different comparative advantages on the margin. For example, a comparative advantage such as different marginal costs of supplying quality leads to a reasonable conjecture that the firms would split the consumer market. However, the effects of competition are more subtle than in the simpler, horizontal-heterogeneity setting with symmetric firms.

Turning to the case of endogenous comparative advantage, we follow [Champsaur and Rochet \(1989\)](#) and consider a two-stage-game duopoly in which firms commit to quality ranges $Q^j = [\underline{q}_j, \bar{q}_j]$, $j = 1, 2$ in the first stage, and then compete using price schedules properly restricted to these intervals in the second stage. Let $\pi^j(Q^1, Q^2)$ be the equilibrium profits in the second stage given the intervals chosen in the first. [Champsaur and Rochet \(1989\)](#) demonstrate that for any given first-stage quality intervals, there is an equilibrium to the second-stage price game with well-defined payoff function $\pi^j(Q^1, Q^2)$, and that in the equilibrium, firms typically find it optimal to leave quality gaps: e.g., $\underline{q}_1 \leq \bar{q}_1 < \underline{q}_2 \leq \bar{q}_2$. In the presence of a gap, firm 1 sells to the lower interval of consumer types with a lower quality product line while firm 2 serves the higher end with higher qualities. More remarkably, they show under a few additional conditions that when a gap in qualities exist, the profits of the firms can be decomposed into two terms:

$$\begin{aligned}\pi^1(Q^1, Q^2) &= \pi^1(\{\bar{q}_1\}, \{\underline{q}_2\}) + \pi^1(Q^1, [\bar{q}_1, \infty)), \\ \pi^2(Q^1, Q^2) &= \pi^2(\{\bar{q}_1\}, \{\underline{q}_2\}) + \pi^2((-\infty, \underline{q}_2], Q^2).\end{aligned}$$

The first term corresponds to the payoffs in a traditional single-quality product-differentiation game. If this was the only component to payoffs, firms would choose their qualities to differentiate themselves on the margin and soften second-stage price competition; this is pure differentiation profit. The second term is independent of the other firm's strategy and represents the surplus extracted from consumers situated entirely in the area of local monopoly; this is pure segmentation profit along the lines of [Mussa and Rosen](#). In many settings, the Chamberlinian incentive to differentiate products in the first term dominates the incentive to segment consumers within the served interval. For example, when preferences are as in [Mussa and Rosen \(1978\)](#), with $u^j(q, \theta) = \theta q$, $C(q) = \frac{1}{2}q^2$ and θ uniformly distributed, [Champsaur and Rochet \(1989\)](#) show that in equilibrium each firm makes a positive profit but offers a unique quality. Although non-linear pricing is an available strategy, both firms optimally discard this option in the first stage to increase profits in the second. Competition acts to reduce choice.

6.2.2. Multidimensional models of heterogeneity

A shortcoming of one-dimensional models is that they are inadequate to capture both privately-known brand preferences and privately-known marginal values of consump-

tion. Furthermore, the lack of compelling conclusions from the one-dimensional modeling approach (particularly the vertical setting) is partly due to the ad hoc manner of modeling product differentiation across firms. Ideally, we would derive economic implications from a model which contains both horizontal and vertical unobservables with a more general allowance for product differentiation.

Two approaches have been taken by the literature. The first relies upon simulations to uncover the basic tendencies of price discrimination.⁶² Borenstein (1985), for example, considers a closely related model with heterogeneity over transportation costs (or “choosiness”) as well as value.⁶³ Using simulations in both second- and third-degree frameworks, he concludes that sorting over brand preferences rather than vertical preferences leads to a greater price differential when markets are very competitive. Borenstein and Rose (1994) develop a similar model to study the impact of competition on price dispersion and conclude from numerical results that dispersion increases for reasonable parameter values.

A second modeling approach to product differentiation uses the well-known discrete-choice approach and incorporates vertical heterogeneity along the lines of Mussa and Rosen (1978). By assuming brand preferences enter utility additively (as in the discrete-choice literature), enough additional structure on preferences arises to provide tractable solutions. In this spirit, Rochet and Stole (2002) develop a methodology for calculating non-linear price schedules when preferences take the discrete-choice form of $u(q_j, \theta) - P_j(q_j) - \xi_j$, and ξ_j represents a brand-specific shock to the consumer when purchasing from firm j .

Each consumer has a multidimensional type given by $(\theta, \xi_1, \dots, \xi_n)$. As before, let $v_j(\theta) \equiv \max_q u(q_j, \theta) - P_j(q)$ represent a type- θ consumer’s indirect utility of purchasing from firm j , excluding brand-specific shocks. When considering a purchase from firm j , the consumer’s outside option is given by

$$\underline{v}_j(\theta, \xi) \equiv \max_{k \neq j} \{0, v_1(\theta) - \xi_1, \dots, v_n(\theta) - \xi_n\} + \xi_j.$$

The competitive environment, from firm j ’s point of view, is entirely contained in the description of $\underline{v}_j(\theta, \xi)$.

First, consider the monopoly case of $n = 1$ and suppose that ξ_1 is independently distributed from θ according to the distribution function $G(\xi)$ on $[0, \infty)$. A monopolist facing a class of consumers with these preferences has a two-dimensional screening problem. The program itself, however, is very easy to conceive. For any non-linear pricing schedule, $P_1(q)$, there is an associated indirect utility, $v_1(\theta) = \max_q u(q, \theta) - P_1(q)$. Because a consumer will purchase if and only if $v_1(\theta) \geq \xi_1$, the monopolist’s market share conditional on θ is $G(v_1(\theta))$. The monopolist’s objective in terms of $q(\theta)$ and $v_1(\theta)$ becomes

$$\max_{\{q, v_1\}} E_\theta [G(v_1(\theta))(u(q(\theta), \theta) - C(q(\theta)) - v_1(\theta))],$$

⁶² E.g., Borenstein (1985), Wilson (1993), among others.

⁶³ Armstrong (2006) provides further analysis in this spirit.

subject to incentive compatibility conditions that $v'_1(\theta) = u_\theta(q(\theta), \theta)$ and $q(\theta)$ is non-decreasing. There is no participation constraint because participation is endogenous. Now $v_1(\theta_0)$ is no longer chosen to equal the outside option of zero (a corner condition in a standard monopoly program), but is instead chosen to satisfy a first-order condition. In a classic tradeoff, a higher utility level (equivalent to shifting $P_1(q)$ downward) reduces the profitability of all inframarginal consumers, but increases market share by raising indirect utilities.

To determine the appropriate first-order conditions of this problem, one needs to appeal to control-theoretic techniques. The resulting Euler equation is a second-order non-linear differential equation with boundary conditions, and generally does not yield a closed-form solution. Nonetheless, Rochet and Stole (2002) demonstrate that in the Mussa and Rosen (1978) model, in the presence of additive utility shocks, the equilibrium quality allocation lies between the first-best solution and Mussa–Rosen solution without additive uncertainty ($\xi_1 \equiv 0$). In this sense, the addition of random effects reduces the monopolist's distortions. While it is true that leaving consumers with more surplus reduces profits for any given rate of participation, raising consumer surplus on the margin has a first-order beneficial effect of increasing the rate of participation. The endogenous participation rate implies it is less profitable to extract surplus from consumers and, therefore, it is no longer as valuable to distort quality in order to extract surplus.⁶⁴

Returning to the setting of competition, we can now analyze models incorporating product differentiation, heterogeneity over brand preferences and heterogeneity over preferences for quality. Several papers have taken this approach in one form or another, including Schmidt-Mohr and Villas-Boas (1999), Verboven (1999), Armstrong and Vickers (2001), and Rochet and Stole (2002). For example, consider a duopoly with two firms on the endpoints of a Hotelling market of unit length, populated with consumers, each with unit transportation costs equal to τ . Let the consumer's location, x , in this interval take on the role of the additive shock. Specifically, $\xi_1 = \tau x$ and $\xi_2 = \tau(1 - x)$, where x is distributed according to some distribution $G(x)$ on $[0, 1]$. As before, firm j makes profit of $u(q(\theta), \theta) - C(q(\theta)) - v_j(\theta)$ for each consumer of type θ who purchases. A consumer will purchase from firm 1, only if $v_1(\theta) - \tau x \geq \max\{0, v_2(\theta) - \tau(1 - x)\}$. Hence, the probability that a consumer of type θ visits the firm $j \neq k$ is given by the participation function:

$$G_j(v_1, v_2) = \begin{cases} G\left(\min\left\{\frac{v_1}{\tau}, \frac{1}{2} + \frac{v_1 - v_2}{2\tau}\right\}\right), & \text{for } j = 1, \\ 1 - G\left(\min\left\{\frac{v_2}{\tau}, \frac{1}{2} + \frac{v_2 - v_1}{2\tau}\right\}\right), & \text{for } j = 2. \end{cases}$$

The two arguments in each of the minimands represent the cases of local monopoly and local competition, respectively.

⁶⁴ Interestingly, these problems may generate a lower interval of pooling, even with the standard restrictions of quadratic preferences and uniformly distributed θ in Mussa and Rosen (1978).

The participation function gives rise to a competitive non-linear pricing game with a well-defined normal form in quality and utility allocations. Each duopolist j chooses the functions (q_j, v_j) to maximize

$$E_{\theta}[G_j(v_1(\theta), v_2(\theta))(u(q_j(\theta), \theta) - C(q_j(\theta)) - v_j(\theta))],$$

subject to the requirement that $v'_j(\theta) = u_{\theta}(q_j(\theta), \theta)$ and $q_j(\theta)$ is non-decreasing. The monopoly methodology of Rochet and Stole (2002) can be directly applied to solve for each firm's best-response function, which in turn can be used to determine the equilibrium price schedules. Generally, closed-form equilibrium solutions are not available, and we must resort to numerical routines to determine solutions. The form of the solution is similar to those of monopoly when the market is not fully covered: the duopoly allocations of quality lie between the case of monopoly and the first-best, with lower marginal prices for quality than the monopolist in Mussa and Rosen (1978). As firms become less differentiated (τ decreases), the duopoly solution converges to the full-information, first-best allocation. Hence, duopoly lies between the monopoly and socially-efficient allocations.

Two final remarks are in order. First, this discrete-choice approach to competition with a single dimension of vertical uncertainty is quite flexible. It can easily be extended to oligopolies with general distributions of ξ . In such an n -firm oligopoly, firm i 's market share is represented by $G_j(v_1, \dots, v_n) \equiv \text{Prob}[u_j - \xi_j \geq \max_{k \neq j} u_k - \xi_k]$, and the analysis proceeds as before. It can also easily be adapted to explore questions about price-cost margins and add-on pricing, as discussed in Section 6.3 below.

Second, and more fundamental, a precise solution for these games can be determined in one instance, as independently noted by Armstrong and Vickers (2001) and Rochet and Stole (2002).⁶⁵ Suppose that the firms are symmetric and that the market is entirely covered in equilibrium. Define the following inverse hazard rate which captures firm j 's ability to profit from its random brand effect as

$$H_j(u_1, \dots, u_n) \equiv \frac{G_j(v_1, \dots, v_n)}{\frac{\partial}{\partial v_j} G_j(u_1, \dots, u_n)}.$$

If this function is homogeneous of degree zero in indirect utilities for each firm (i.e., $\frac{d}{dv} H_j(v, \dots, v) = 0$ for each j), then non-linear, cost-plus-fixed-fee prices, $P_j(q) = C(q) + F_j$, form a Nash equilibrium. Such homogeneity naturally arises when the fixed-effects distributions are symmetric. For example, in our Hotelling duopoly above, $H_j(v, v) = \tau$ and the equilibrium prices are $P(q) = C(q) + \tau$. The similarity with the uniform-pricing, single-product Hotelling game is remarkable. Duopolists earn profits over their locational advantage, but because they have no competitive advantage in supplying quality, they do not gain from distortions on this margin. This result of cost-plus-fixed-fee pricing, however, depends critically upon firm symmetries and upon market coverage. Changes in either of these assumptions will open up the possibility that firms distort qualities in their battle for market share.

⁶⁵ Verboven (1999) finds a related pricing result: given exogenous quality levels and cost symmetries between duopolists, price is equal to cost plus a uniform markup.

6.3. Applications: add-on pricing and the nature of price–cost margins

Several papers use multidimensional discrete-choice frameworks to explore specific price discrimination questions. We consider here the contributions in Verboven (1999) and Ellison (2005).

Verboven (1999) uses this framework to make predictions about absolute and relative price–cost margins, and how these margins move with respect to quality. According to the received theory of monopoly non-linear pricing, the monopolist’s absolute price–cost margins, $P(q) - C(q)$, increase with quality, *but* the percentage price–cost margins, $(P(q) - C(q))/P(q)$, typically *fall* with q . This finding is certainly true for the Mussa–Rosen (1978) setting in Section 6.1, and it is also true for a monopolist choosing two qualities and selling to two types of consumers, $\theta \in \{\underline{\theta}, \bar{\theta}\}$.⁶⁶ However, this theoretical prediction seems at odds with reality. For example, Verboven (1999) presents evidence from the European automobiles that the margins on quality add-ons are higher than the margins on base products. This leads one to reject the simple monopolistic model of second-degree price discrimination and conclude that the percentage price–cost margins rise with quality for this market.

In response, Verboven (1999) makes two changes to the basic monopoly model to create an alternative theory that better fits the data and provides a remarkably simple model of competition with non-linear pricing. First, assume that consumers have both a vertical heterogeneity component, θ , and a horizontal fixed effect for each product line. Second, assume that high-quality prices are unobservable by consumers. For example, it may be that prices for quality add-ons are not advertised and would be costly to learn before arriving at the point of sale. Ellison (2005) refers to a setting with this second assumption as an “add-on pricing” game.

The first modification requires distributional assumptions. Rather than following Verboven (1999), we use a simpler set of distributional assumptions to the same effect. We assume that θ takes on only two equally likely types, $\theta \in \{\underline{\theta}, \bar{\theta}\}$ with $\bar{\theta} > \underline{\theta}$, and that brand preferences derive from a Hotelling model of differentiation; i.e., the firms are positioned on the endpoints of a Hotelling market of length 1, consumers are uniformly distributed and must expend transportation costs of τ per unit-distance traveled.⁶⁷ The firms can sell only two exogenously given qualities, q_1 and q_2 with $\Delta q = (q_2 - q_1) > 0$, at costs of $c_2 \geq c_1$, respectively. For the high type of consumer, we also suppose the efficiency of a quality add-on: $\bar{\theta} \Delta q > c_2 - c_1$. Applying the results of Armstrong and Vickers (2001) and Rochet and Stole (2002), if the market is covered

⁶⁶ Note that this result for the two-type case relies upon the monopolist optimally choosing qualities with smooth, convex cost function. For arbitrary qualities, it is no longer necessarily true. In Verboven (1999), the underlying type distribution for θ is continuous and only two exogenous qualities are offered. Here, the result of decreasing relative price–cost margins again emerges.

⁶⁷ These different assumptions do not change Verboven’s main theoretical conclusions and allow us to more closely compare Verboven (1999) to recent work by Ellison (2005).

and consumers observe the full price schedules, then equilibrium prices are cost-plus-fixed-fee. It follows that equilibrium prices under duopoly with fully advertised price schedules are $p_2 = c_2 + \tau$ and $p_1 = c_1 + \tau$, and incentive compatibility constraints are non-binding. Absolute price–cost margins are constant but, as in monopoly, percentage price–cost margins fall with quality.

To this duopoly model, Verboven (1999) adds a second modification, similar to that in Lal and Matutes (1994): while consumers costlessly observe the base-quality price, they cannot observe the high-quality price without expending an additional search cost. Consumers correctly anticipate the high-quality product prices and in equilibrium have no incentive to search, regardless of the insignificance of the search cost. As a result, the individual firms behave as monopolists on the high-quality item. Under some reasonable conditions for types and costs, the firms will set a price that makes the high-type consumer indifferent between consuming the high- and low-quality products: $p_2 = p_1 + \bar{\theta} \Delta q$. The low-quality price now serves a new role: it provides a credible signal about the unobserved high-quality price. Each firm, taking the equilibrium prices of its rival as given, $\{p_1^*, p_2^*\}$, chooses its low-quality price to maximize the following expression (after substitution of $p_2 = p_1 + \bar{\theta} \Delta q$ and $p_2^* = p_1^* + \bar{\theta} \Delta q$):

$$(p_1 - c_1) \left(\frac{1}{2} + \frac{p_1^* - p_1}{2\tau} \right) + (p_1 + \bar{\theta} \Delta q - c_2) \left(\frac{1}{2} + \frac{p_1^* - p_1}{2\tau} \right).$$

Solving for the optimal price (and imposing symmetry), we have $p_1^* = \bar{c} + \tau - \frac{\bar{\theta}}{2} \Delta q$ and $p_2^* = \bar{c} + \tau + \frac{\bar{\theta}}{2} \Delta q$, where $\bar{c} \equiv (c_1 + c_2)/2$.

Two interesting results follow when there is incomplete information about add-on prices. First, relative price–cost margins increase with quality if and only if competition is sufficiently strong, $\tau < (\bar{\theta} \Delta q - \Delta c) \bar{c} / \Delta c$. This result is in sharp contrast to the monopoly and fully-advertised duopoly pricing games. Second, although Verboven (1999) does not stress this result, the presence of incomplete information about add-on prices does not raise industry profits; instead, average prices are unchanged. Indeed, for any fixed price differential, $p_2 - p_1$, the first-order condition for profit maximization entirely pins down the average price; the average price equals the average cost plus τ . The ability to act as a monopolist on the high-quality item gets competed away on the low-type consumers; the result is similar to Lal and Matutes' (1994) for loss-leader pricing.

It is perhaps surprising that the presence of incomplete information about high-quality prices does not increase profits, especially given that the reasoning in Diamond (1971) suggests unobservable pricing generates monopoly power in other contexts.

In order to study the strategic effects of unobservable add-on pricing further, Ellison (2005) uses a similar model to the one presented above but with the key difference of heterogeneity, γ , over the marginal utility of income: $u = q - \gamma p - \tau x$, where x is the distance traveled to the firm. Defining $\theta = 1/\gamma$, these preferences can be normalized to $u(q, \theta) = \theta q - p - \theta \tau x$. While this formulation has the same marginal rate of substitution between money and quality as Verboven's (1999) setting, a consumer with

higher marginal utility of income (and therefore a lower marginal willingness to pay for quality) now *also* is less sensitive to distance. This vertical heterogeneity parameter captures price sensitivities with respect to a competitor's price and to a firm's own price. This reasonable modification of preferences also eliminates Verboven's (1999) profit neutrality result.

To see why this matters, we can introduce two distinct terms, τ_1 and τ_2 , to capture the sensitivities of each market.⁶⁸ The first-order condition requires

$$\left(\frac{1}{2} - \frac{1}{2\tau_1}(p_1 - c_1)\right) + \left(\frac{1}{2} - \frac{1}{2\tau_2}(p_2 - c_2)\right) = 0.$$

Given the unobservability of add-on prices, profit maximization and incentive compatibility require a positive price differential, $p_2 - p_1 = \bar{\theta}\Delta q > 0$, and the marginal effect of a price change will be positive in market 1 and negative in market 2. In Verboven's (1999) setting, $\tau_1 = \tau_2 = \tau$, because there is no heterogeneity over brand sensitivities. The effect on profits from a marginal reduction in p_1 is equal to the effect of a marginal increase of p_2 . It is optimal that the average price is unchanged and that the individual prices are equally distorted from $\bar{c} + \tau$. In contrast, after normalizing utility, Ellison's (2005) setting has $\tau_1 = \underline{\theta}\tau$ and $\tau_2 = \bar{\theta}\tau > \tau_1$. Now, a small reduction in p_1 has a greater effect on profit than an equal increase in p_2 does. Hence, the base price is less downward distorted than the high-quality price is upward distorted. The price dispersion is not centered around $\bar{c} + \tau$, and the net result is an increase in the average price (and profit) from unobservable add-on prices [Ellison (2005)]. While the profit neutrality result in Verboven (1999) is quite interesting, it should be applied cautiously given the plausibility of the preferences in Ellison (2005).

6.4. Non-linear pricing with consumers in common

In the previous section, the models were cast with the discrete-choice, one-stop-shopping assumption that assigns each consumer to at most one firm in equilibrium. The only conduit for competitive effects was through the outside option; once a consumer chose to purchase from a firm, the offers of other firms became irrelevant. This assumption generated a natural separability in the equilibrium analysis. In some settings, however, one-stop-shopping is an inappropriate assumption. As an extreme example, one could imagine two firms selling complementary goods such as a monopoly vendor of software and a monopoly vendor of computer hardware. In a less extreme example, two firms may sell differentiated products that are substitutes, but it is efficient for the customer to consume some output from each seller. In both cases, the consumer is likely to purchase from both firms in equilibrium, and the non-linear price schedule offered by one firm will typically introduce a competitive externality on its rival's sales *at every margin*.

⁶⁸ We assume the sensitivities are not too different from one another.

6.4.1. One-dimensional models

Most common agency models consider competition between two principals (e.g., price-discriminating duopolists) for a common agent's activities (e.g., consumer's purchases) when there is one dimension of uncertainty over the agent's preferences.⁶⁹ In equilibrium, the consumer purchases goods from *both* firms, which introduces a new difficulty: one firm's contract can negatively impact the screening effectiveness of the other's.⁷⁰

The most interesting and tractable setting for such one-dimensional models are when (i) both firms care about the same dimension of preference uncertainty, and (ii) consumption of one firm's good affects the marginal utility of consuming the other firm's good. An example of the first condition arises when the consumer's marginal utility of income (e.g., high marginal utilities of income may imply high price elasticities of demand for all goods) is relevant. An example of the second condition arises when the goods are either substitutes (i.e., $u_{q_1 q_2}(q_1, q_2, \theta) < 0$) or complements (i.e., $u_{q_1 q_2}(q_1, q_2, \theta) > 0$). If there is no interaction in the consumer's utility function, then the firms are effectively monopolists over their products and competition is not economically meaningful.

In the common-agency game, each firm j simultaneously offers the consumer a non-linear price schedule, $P_j(q_j)$, for the purchase of good q_j . The consumer decides how much (if any) he wishes to buy of the two goods, and then makes his purchases simultaneously. Importantly, firms cannot condition their price schedule on the consumer's choices from the rival.

Suppose that the consumer's utility is quasi-linear and characterized by

$$u(q_1, q_2, \theta) = P_1(q_1) + P_2(q_2),$$

⁶⁹ Stole (1991), Martimort (1992), and Martimort and Stole (2005) present the basic analysis. Other related papers, covering a variety of applications, include Gal-Or (1991), Laffont and Tirole (1991), Biglaiser and Mezzetti (1993), Bond and Gresik (1996), Martimort (1996), Mezzetti (1997), Calzolari (2004), and Martimort and Stole (2003). An early paper by Calem and Spulber (1984) considers a common agency setting in which firms are restricted to offering two-part tariffs.

⁷⁰ For completeness, one needs to distinguish between *intrinsic* and *delegated* common agency games. This distinction was first noted by Bernheim and Whinston (1986) in the context of common agency and moral hazard. When the common agency game is intrinsic, the agent is assumed to be unable to accept only one of the two principals' offers. This is an appropriate assumption, for example, in regulatory settings where the regulated firm can either submit to all governmental regulatory bodies or exit the industry. When common agency is delegated, the agent has the additional options of contracting with just one or the other principal. When firms cannot monitor a consumer's purchases with a rival, the delegated common agency game is more appropriate. As noted by Martimort and Stole (2005), however, the distinction has no impact on the equilibrium allocations of $q_j(\theta)$ chosen by participating consumers at each firm. The distinction does matter if one is interested in consumer surplus (which is higher under delegated agency) or market participation when coverage is incomplete (delegated agency games may generate more market coverage than intrinsic agency games). Of course, when the goods are perfect complements on the extensive margin, such as arguably in the example of monopoly computer software and hardware firms, the games are strategically equivalent since a consumer would never choose to purchase from only one firm.

which satisfies a one-dimensional single-crossing property in each (q_j, θ) pair and is increasing in θ . Appealing to previous arguments, if q_2 was fixed, firm 1 would construct a non-linear pricing schedule to induce consumers of type θ to select $q_1(\theta)$, satisfying the relationship

$$u_{q_1}(q_1(\theta), q_2, \theta) - C'(q_1(\theta)) = \frac{1 - F(\theta)}{f(\theta)} u_{q_1\theta}(q_1(\theta), q_2, \theta).$$

Generally, however, when $u_{q_1q_2} \neq 0$, the choice of q_2 will depend upon the offer of $P_1(q)$.

We proceed, as before, by converting the competitive problem into the more familiar and tractable monopoly program. To this end, take firm 2's pricing schedule as fixed and define the consumer's best-response function and indirect utility, given (q_1, θ) :

$$\hat{q}_2(q_1, \theta) = \arg \max_{q_2} u(q_1, q_2, \theta) - P_2(q_2).$$

$$v_1(q_1, \theta) = \max_{q_2} u(q_1, q_2, \theta) - P_2(q_2).$$

The indirect utility function, $v_1(q, \theta)$, is continuous and increasing in both arguments. It is straightforward to check that if the goods are complements, then $v_1(q, \theta)$ satisfies the single-crossing property.⁷¹ If the goods are substitutes, then whether single-crossing is satisfied must be checked in equilibrium. For a wide variety of preferences, this concern is not a problem so we cautiously proceed by setting it aside. The end result is that firm 1's optimization problem is identical to that of a monopolist facing a consumer with preferences $v_1(q, \theta)$.⁷² Competitive effects are embedded in this indirect utility function, much as they are embedded in the outside option $\underline{v}_j(\theta)$ when there is one-stop shopping.

Suppose for the sake of argument that \hat{q}_2 is continuous and differentiable and $v_1(q, \theta)$ satisfies the single-crossing property. Then, using the monopoly methodology, firm 1's optimal price-discriminating solution is to choose $q_1(\theta)$ to satisfy

$$\frac{\partial v_1(q_1(\theta), \theta)}{\partial q_1} - C_q(q_1(\theta)) = \frac{1 - F(\theta)}{f(\theta)} \frac{\partial^2 v_1}{\partial q_1 \partial \theta}(q_1(\theta), \theta).$$

Using the equilibrium condition that $q_2(\theta) = \hat{q}_2(q_1(\theta), \theta)$ and applying the envelope theorem to replace the derivatives of v_1 with derivatives of \hat{q}_2 , we obtain pointwise in θ :

$$\begin{aligned} & u_{q_1}(q_1, q_2, \theta) - C_q(q_1) \\ &= \frac{1 - F(\theta)}{f(\theta)} \left(u_{q_1\theta}(q_1, q_2, \theta) + u_{q_1q_2}(q_1, q_2, \theta) \frac{\partial \hat{q}_2(q_1, \theta)}{\partial \theta} \right). \end{aligned} \quad (2)$$

⁷¹ Formally, complementarity and single-crossing implies that $u(q_1, q_2, \theta) - P_1(q_1) - P_2(q_2)$ is supermodular in (q_1, q_2, θ) . Hence, the maximized function is also supermodular.

⁷² A few technical issues regarding the associated virtual surplus function – namely strict quasi-concavity and supermodularity – must also be addressed; see Martimort and Stole (2005), for details.

Comparing this result to the analogous monopoly equation, (1), we see that the presence of a duopolist introduces a second information-rent effect: $u_{q_1 q_2} \frac{\partial \hat{q}_2}{\partial \theta}$.

Suppose for the moment that the duopolists' goods are substitutes: $u_{q_1 q_2} < 0$. Because \hat{q}_2 is increasing in θ , the second term is negative and reduces the standard distortion. Hence, when the products are substitutes in the consumer's preferences, distortions are reduced by competition. Alternatively, if the goods were complements, the distortions would be amplified. Intuitively, selling an extra margin of output to a lower type requires that the firm reduce the marginal price of output for all consumers (including higher types), and in so doing, reduce inframarginal profit. The reduction in inframarginal profit, however, is offset by the fact that a marginal increase in q_1 causes the consumer to lower q_2 marginally, which consequently lowers the information-rent term u_θ .

In a broader sense, the result shares the same spirit as Bertrand equilibria in pricing games with differentiated, single-product duopolists. If the goods are imperfect substitutes, we know the presence of competition reduces purchasing distortions, while if the goods are complements, distortions increase. This intuition goes back at least as far as Cournot (1838). The present argument suggests that this single-price intuition is robust to the introduction of more complicated non-linear pricing and multi-product firms.

6.4.2. Multidimensional models

One can easily think of examples in which the two-dimensional preference uncertainty is more appropriate because each firm wants to segment on a different dimension of consumer tastes. In these settings, the interaction of marginal utilities of consumption may introduce a common agency setting worthy of study. Unfortunately, this approach shares many of the same technical difficulties with multidimensional screening models and so has received little attention. A notable exception is the paper by Ivaldi and Martimort (1994).⁷³ The authors construct a model which avoids many of these technical difficulties by employing a clever change of variables to allow aggregation of consumer heterogeneity into a one-dimensional statistic.⁷⁴

A simple version of Ivaldi and Martimort's (1994) model makes this technique clear. Suppose that there are two competing firms, $i = 1, 2$, each producing one good and offering non-linear pricing schedules $P_i(q_i)$ to the population of consumers. A consumer has two-dimensional private information, (θ_1, θ_2) , and preferences for consumption of the two goods given by

$$u = \theta_1 q_1 + \theta_2 q_2 - \frac{1}{2} q_1^2 - \frac{1}{2} q_2^2 + \lambda q_1 q_2 - P_1 - P_2,$$

⁷³ Miravete and Röller (2003) successfully apply a similar methodology to their study of the U.S. cellular telephone industry. A related competitive setting in which aggregation usefully converts a multidimensional problem into a single dimension can be found in Biais, Martimort and Rochet (2000).

⁷⁴ Ivaldi and Martimort (1994) empirically fit this model to data from the French energy market.

with $|\lambda| < 1$. For the moment, suppose firms are restricted to offering quadratic price schedules. Taking the price schedule of firm 2 as given, $P_2(q_2) = \alpha_2 + \beta_2 q_2 + \frac{\gamma_2}{2} q_2^2$, it follows that the type (θ_1, θ_2) consumer's first-order condition for choice of q_2 is given by

$$\theta_2 - q_2 + \lambda q_1 = \beta_2 + \gamma_2 q_2.$$

Solving for q_2 and substituting in the first-order condition for the choice of q_1 , yields

$$\theta_1 - q_1 + \frac{\lambda}{1 + \gamma_2} (\theta_2 - \beta_2 + \lambda q_1) = P_1'(q_1).$$

We can define a new measure of heterogeneity as $z_1 \equiv \theta_1 + \frac{\lambda \theta_2}{1 + \gamma_2}$, and use it as a one-dimensional sufficient statistic for firm 1's consumer preferences. The two-dimensional problem has thus been simplified and standard methods can be employed. Furthermore, providing that z_1 follows a Beta distribution with parameter λ , Ivaldi and Martimort (1994) show that firm 1's optimal contract is indeed quadratic.⁷⁵ The conclusions of the model are quite simple: an increase in λ (tantamount to making the goods closer substitutes) causes the duopolists to reduce price margins, further suggesting that our intuition from the differentiated Bertrand model is robust to multidimensional types and to larger strategy spaces which include non-linear price schedules.

7. Bundling

It is well known that a multiproduct monopolist can increase its profit by engaging in some form of bundling, even when demands for the component products are independently distributed.⁷⁶ The intuition of monopoly bundling is well known. Take a monopolist selling two distinct goods, for example, who can offer them for sale individually or as a bundle, with the bundle price being less than the individual prices. Such pricing effectively segments the market into three groups: those with moderately high valuations for both goods who buy the bundle, those with high valuations for one good and low valuations for the other who buy at the individual prices, and those who do not purchase. One can think of this pricing strategy as a form of non-linear pricing with quantity discounts: the first unit is for sale at the individual price, and the second unit is for sale at a reduced price equal to the difference between the bundle price and the sum

⁷⁵ While this condition places rather strong restrictions on the equilibrium distribution of (θ_1, θ_2) , one could utilize the simple aggregation approach for more general distributions, providing one was content to restrict strategy spaces to quadratic price schedules.

⁷⁶ McAfee, McMillan and Whinston (1989) provide a general set of results in this context. See also Stigler (1963), Adams and Yellen (1976), Schmalensee (1984), Armstrong (1999), Armstrong and Rochet (1999) and Bakos and Brynjolfsson (1999) for more on monopoly bundling strategies. For a discussion of the same effect in the context of regulation, see Section 4.5 of Armstrong and Sappington (2007) in this volume.

of the individual prices.⁷⁷ Moreover, if consumer reservation values are independently distributed across goods, as the number of goods increases, the law of large numbers implies a homogenizing effect as it reduces consumer heterogeneity. Mixed bundling with a large number of goods allows the monopolist to extract almost all of the consumer surplus, as shown by Armstrong (1999).⁷⁸

When bundling is introduced in markets with imperfect competition, additional effects arise. In what follows, there are three pricing strategies to evaluate. A firm can offer “components pricing” in which each product is sold for a separate price and there is no discount for multiple purchases from a single firm; a firm can practice “pure bundling” (or “tying”) by offering only a package of goods for a bundled price; and a firm can practice “mixed bundling” by offering both component prices and bundle discounts. It is also helpful to proceed by separately examining these strategies in the context of multiproduct duopoly and in the context of multiproduct monopoly with the threat of entry by single-product firms.

7.1. Multiproduct duopoly with complementary components

Suppose that there are two firms in an industry, a and b , each producing in two distinct markets, 1 and 2. We assume, however, that the two component goods are perfect complements. For example, market 1 consists of two brands of stereo receiver and market 2 consists of two brands of speakers. Assuming that the goods are compatible, consumers have preferences over the four possible combinations of goods that make up a system.⁷⁹

Consider our three marketing strategies: components pricing, pure bundling, and mixed bundling. In the first strategy, firm a offers the pair of component prices $\{p_{1a}, p_{2a}\}$; under pure bundling, firm a offers a single price, p_{12a} ; and under mixed bundling, firm a offers three prices, $\{p_{1a}, p_{2a}, p_{12a}\}$. Absent the ability to publicly commit to an overall marketing strategy before prices are chosen, each firm individually prefers to choose mixed bundling for reasons familiar to a monopolist. Such a choice, however, may lead to lower industry profits than had the firms chosen component pricing. Matutes and Regibeau (1992) make this point as do Anderson and Leruth (1993) for different demand structures.⁸⁰ That industry profits may be lower is reminiscent of

⁷⁷ Formally, optimal pricing in these environments quickly becomes intractable because sorting occurs over a multidimensional space. See Armstrong and Rochet (1999) and Rochet and Stole (2003) for a survey of the multidimensional screening literature and the results on bundling by a monopolist.

⁷⁸ When marginal costs are zero, there is no social loss from selling consumers only pure bundles, as valuations will exceed marginal costs on every component. Hence, for information goods which typically have very low marginal costs, the homogenizing effect of pure bundling is especially attractive. Bakos and Brynjolfsson (1999, 2000) emphasize this point.

⁷⁹ The issue of the choice of compatibility is addressed in Matutes and Regibeau (1988, 1992) and Economides (1989).

⁸⁰ Matutes and Regibeau (1992), in a model of Hotelling product differentiation, show that when the components are compatible, mixed bundling by both duopolists generates lower industry profits than pure component

Thisse and Vives (1988) and other asymmetric best-response models in which price discrimination causes intense competition. In mixed-bundling case, however, there is no clear notion of weak and strong market asymmetries, so the intuition for why mixed bundling lowers profits relative to the other strategies is not immediate. To understand, suppose that a firm switches from pure components pricing to mixed-bundling. Such a firm will increase its profits (holding the prices of the rival fixed) by choosing its bundle price to be than lower the total of its pure component prices and its new component prices to be higher than before. The result is that the rival's residual demand for mixed-systems will fall. If the rival anticipates the other firm's mixed bundling strategy, it will lower its component prices. In the end, mixed bundling reduces industry profits on net. Of course, if the market is covered, mixed bundling also reduces social welfare relative to pure component pricing.

The previous analysis suggests that firms may have an incentive to commit to either pure bundling or pure components pricing in order to avoid the intense competition of mixed bundling. Matutes and Regibeau (1988) and Economides (1989) take up this issue in a dynamic game which allows firms to commit to either pure bundling or pure components pricing prior to choosing prices. They find that pure components pricing raises industry profits relative to pure bundling, and that it is a dominant strategy equilibrium for each firm to choose component pricing.

The economic intuition for their result is simple. Suppose under components pricing that a firm a reduces its price for product 1 by Δp . The result is that the demand for firm a 's market-1 good increases without any effect in market 2. Under pure bundling, if firm a reduces its system price by Δp , the demands for its goods in both markets increase. In short, demand is more elastic when firms practice pure bundling. Profits are lower as a result.

Matutes and Regibeau (1992) and Anderson and Leruth (1993) generalize the game to allow for a commitment to mixed-bundling, in addition to allowing pure-components and pure-bundling strategic commitments. A similar economic intuition is present in both papers; mixed bundling is unprofitable from an industry perspective, and a commitment to pure component pricing may arise in equilibrium. Matutes and Regibeau (1992) find that depending upon costs, it is an equilibrium for firms to commit to either pure component or mixed bundling strategies in the initial stage.⁸¹ In a different model of demand, Anderson and Leruth (1993) find that a commitment to pure-component pricing always arises in equilibrium, and industry profits are higher than if commitment does not exist and mixed-bundling occurs.

pricing. Anderson and Leruth (1993), find a similar relationship when preferences for differentiation are drawn from Logit distributions.

⁸¹ When costs are low, a "prisoner's dilemma" exists between the strategies of mixed bundling and components pricing, resulting in firms committing to mixed bundling. For intermediate costs, one firm chooses mixed bundling, the other chooses component pricing, and for high costs both firms choose component pricing.

7.2. Multiproduct monopoly facing single-product entry

Consider the setting of an incumbent monopolist facing the threat of entry in one of its two markets. What are the incumbent's incentives to offer bundled prices? Does the answer depend upon the incumbent firm's ability to commit?

One of the first in-depth studies of the value of a commitment to pure bundling ("tying") is found in Whinston (1990).⁸² Whinston addresses the question of whether tying its market-1 good with a product sold in a second (potentially duopoly) market is a profitable strategy to maintain monopoly in both markets. The classical response to this question has been that if the second market exhibits perfect competition and marginal cost pricing, such bundling cannot be valuable to the monopolist. Moreover, profits are reduced if there are some bundle-purchasing consumers whose value of good 2 is lower than the marginal cost of production.

Whinston's model departs from the classical argument by assuming that competition in market 2 is imperfect. In the first model of Whinston (1990, Section I), the monopolist's commitment to bundle its products in two markets lowers the profits of any duopolist who may enter market 2. Because the consumer's valuation of the monopolist's market-1 good is assumed to be independent of his market-2 valuation, and because the monopolist's market-2 good is an imperfect substitute for the entrant's good, tying causes the monopolist to price aggressively following entry. A commitment to sell only a pure bundle makes the monopolist more aggressive in market 2. Under bundling, each additional market share has an added value equal to the profit margin created in market 1. Since the value of a marginal market-2 sale is higher when products are tied, the incumbent prices more aggressively and lowers the profits of firm *b*. For a range of entry costs, the potential entrant will remain out of the market if and only if the monopolist in market 1 makes such a commitment. Of course, such a commitment carries a corresponding cost to the incumbent. Having succeeded in foreclosing entry, the monopolist is now left to maximize profits over the two markets using only a pure bundle price.

A similar entry-deterring effect of tying is explored in Choi and Stefanadis (2001). Suppose that the goods in the two markets are perfect complements, and that a potential entrant exists for each market. An entrant competes in a market only if it has an innovation that makes its component cheaper to produce than the monopolist's. If both entrants succeed in innovating, the monopolist is displaced and earns zero profits. If only one succeeds, however, the monopolist is able to practice a prize squeeze and extract some of the entrant's social contribution, providing its products are not tied. Suppose that the monopolist can make a commitment to tie its products before entry is decided, and that the probability of an innovation depends upon the investment by the potential entrant. Then the monopolist will tie its goods if its ability to extract rents in a price squeeze is

⁸² Carlton and Waldman (2002) develop and extend the ideas in Whinston (1990) to understand how bundling complementary products can be used to preserve and create monopoly positions in a dynamic setting.

sufficiently poor. The tradeoff is a simple one: tying reduces the incentives to invest in innovation and entry, but if a single firm enters, tying prevents the monopolist from squeezing the entrant.

The above examples illustrate how a commitment to pure bundling may deter entry. If entry occurs, of course, the incumbent would like to abandon its tying strategy so credible commitment is critical. As the papers on multiproduct duopoly and mixed bundling suggest, however, an incumbent firm without commitment ability would choose mixed bundling as a statically optimal strategy, and this choice may reduce duopoly profits sufficiently to deter entry. Whinston (1990, Section II(B)) considers a variant of his earlier model with a heterogeneous captive market, showing the possibility that bundling will be ex post profitable with entry, so commitment is unnecessary. In this setting, bundling may sufficiently reduce the profitability of market 2 for the potential duopolist and deter entry. A similar idea of entry deterrence also appears in the discussion of Matutes and Regibeau (1992) in the context of mixed bundling. Because mixed bundling lowers industry profits relative to pure components pricing, a prohibition against price discounts for bundles would raise the likelihood of entry.⁸³

Bundling impacts not only the entry decision, but it may also affect the post-entry profitability of an incumbent. Consider the issue of entry accommodation. Whinston (1990, Section II(A)) makes the point that the strategy of pure bundling will be useful in softening competition following entry. As a simple example, one can imagine intense competition on market 2 without bundling (because the firms' market-2 goods are homogeneous). Now bundling may generate a form of vertical differentiation with the bundled product representing a higher "quality" product with additional features. With such vertical differentiation, pricing in market 2 may be less intense. In this way, a commitment to pure bundling may be an ideal accommodation strategy.

Other authors have found similar accommodating effects from commitments to sell a pure bundle. Carbajo, de Meza and Seidmann (1990) construct a model in which goods are homogeneous in market 2, but values for the goods in market 1 and 2 are perfectly correlated. Pure bundling segments the market into high-valuation consumers purchasing the bundle and low valuation consumers purchasing the single product from firm *b*. If the cost of production in market 2 is not too much greater than in market 1 and demand is linear, then firm *a* will commit to bundling, prices will rise in market 2 and consumer surplus will decrease. Chen (1997a) presents a similar model in which duopolists who can produce in both markets play a first-stage game over whether to sell good 1 or a pure bundle of goods 1 and 2; as before, market 2 is a perfectly competitive market. The pure-strategy equilibrium exhibits differentiation with one firm choosing good 1 and the other offering the pure bundle. Again, bundling serves to soften competition by introducing product differentiation. Importantly, the product differentiation

⁸³ The papers of Nalebuff (2004) and Bakos and Brynjolfsson (2000) demonstrate that pure bundling may be more profitable than component pricing for a multiproduct incumbent, but pure bundling reduces a single-market entrant's profitability. Mixed bundling is not considered.

role of bundling in these models arises because firms commit to sell only the pure bundle; mixed bundling would undermine product differentiation.⁸⁴

8. Demand uncertainty and price rigidities

The pricing strategies studied in the preceding sections are well-known forms of price discrimination. In the present section, we consider a setting which on the surface is less related to price discrimination – specifically, aggregate demand uncertainty with non-contingent (or rigid) pricing.

Imagine that a firm is facing a continuum of heterogeneous consumers, each with unit demands, and that these demands have a common component of uncertainty. For any given realization of the aggregate uncertainty, the market demand curve is downward sloping. We will use $D_s(q)$ to indicate the market demand curve in state s , where $s = 1, \dots, S$ and the probability of a state occurring is given by f_s , $\sum_s f_s = 1$. The firm cannot discriminate over the interconsumer heterogeneity because of the unit demand assumption. The firm may, however, be able to set its uniform price conditional on the realized demand state. With such contingent pricing, the firm would be acting like a third-degree price discriminator where each market is associated with a different demand state. The analysis would mirror that in Section 3.

Suppose instead that firms cannot practice contingent pricing and must fix their prices prior to the realization of aggregate demand. Although firms cannot directly price discriminate by state, they can accomplish some discrimination indirectly by offering buckets of output at various prices. For example, in a low demand state, only the low priced output is purchased. In a high demand state, high-priced output is purchased after the low-priced output is sold out. A higher average price results in the higher demand states. In a sense, firms are indirectly price discriminating across states of aggregate demand by offering fixed amounts of output at a variety of prices.

To focus our discussion, consider our non-contingent pricing assumption applied to a market of homogeneous goods in which aggregate demand uncertainty exists. Following the approach taken by Prescott (1975), Eden (1990) and Dana (1998, 1999a, 1999b), we assume that each firm $i = 1, \dots, n$ offers a distribution of prices and quantities, $q_i(p)$, where $q_i(p)$ gives the number of units available from firm i at price p . We denote the cumulative supply function (i.e., the total amount of output supplied at a price equal to or less than p) as $Q_i(p)$. Let $q(p) = \sum_{i=1}^n q_i(p)$ and $Q(p) = \sum_{i=1}^n Q_i(p)$.

Note that the output-distribution strategy $q_i(p)$ is *not* a non-linear price schedule in the sense of Section 6. In fact, $Q_i(p)$ represents something closer to a supply function as it gives the total number of units supplied by firm i at prices no greater than p . Even so, $Q_i(p)$ is *not* a supply function in the neoclassical sense either, because each firm

⁸⁴ Reisinger (2003) presents a demand model that allows general correlations of brand preferences across products to determine when bundling will soften or intensify competition.

sells its output at a variety of prices along the curve $Q_i(p)$, and typically some rationing occurs as too many customers line up to purchase the lowest priced items.

The firms anticipate the effects of demand uncertainty and offer a distribution of output at different prices. If demand is unusually low, a small measure of consumers shows up and purchases the cheapest priced units; if demand is unusually large, many consumers show up (first buying up the cheap items, then the ones more dear) eventually driving the price of each firm's goods upward.⁸⁵ The greater the positive demand shock, the larger the average selling price will be.

The market for airline tickets is a good example of this phenomenon. In their revenue management programs, airlines try to accomplish several objectives – properly price the shadow costs of seats and planes, effectively segment the market using a host of price discrimination devices (e.g., Saturday night stay-over restrictions, etc.), and quickly respond to changes in aggregate demand. This latter objective is accomplished by offering buckets of seats at different prices. Thus, if a convention in Chicago increases demand for airline tickets on a given weekend from New York to Chicago, the low-priced, restricted-fare economy buckets quickly run dry, forcing consumers to purchase otherwise identical seats at higher prices.⁸⁶

In the demand uncertainty setting with ex post price rigidities, multiple prices are offered in any given aggregate demand state. Therefore, we need a rationing rule for allocating goods to consumers before proceeding. Three different assumptions of rationing have been considered in this literature: proportional, efficient and inefficient. Proportional rationing requires that all consumers willing to buy at a given price are equally likely to obtain the limited quantities of the good. Because we have assumed

⁸⁵ Alternatively, one could model demand uncertainty in a setting in which each customer selects a firm knowing that stock-outs occur with some probability and the customer has no recourse to visit another firm. In these settings, firms can compete on availability through their reputation and pricing strategies. See, for example, the papers by Carlton (1978), Deneckere and Peck (1995) and Dana (2001b), in which firms compete in both price and availability. Because these settings are further removed from the methodology of price discrimination, we do not discuss them here.

⁸⁶ Two related sets of work should also be mentioned and distinguished before proceeding. The first concerns the optimal flexible price mechanism for responding to demand uncertainty. For example, an electrical utility may be capacity constrained in periods of high demand, but it can sell priority contracts to end users to allocate efficiently the scarce output in such periods. Wilson (1993, chs. 10–11) provides a detailed survey of this literature. To my knowledge, no work has been done which examines the effect of competition on such priority mechanisms, although some of the results from the literature on competitive non-linear pricing would be relevant. The second related set of models analyzes supply-function games and equilibria. In these games, firms submit neoclassical supply functions and a single price clears the market in each demand state. The most influential paper on such supply-function equilibria is by Klemperer and Meyer (1989) [see also their earlier paper, Klemperer and Meyer (1986)]. The fundamental difference between supply-function games and the output-distribution games which are the focus of the present section is that, in the former, a Walrasian auctioneer determines the unique market clearing price, while in the latter, a set of prices is available and rationing exists at all but the highest chosen price. Although in both settings average price increases with demand, in a supply-function equilibrium there is no price variation within any aggregate demand state. To preserve focus, we leave this interesting literature aside.

that our consumers have unit demands, this is equivalent to assuming that consumers arrive in random order and purchase at the lowest available price. Efficient rationing assumes that the highest value customers show up first.⁸⁷ Finally, inefficient rationing assumes that those with the lowest valuation for consumption arrive first (perhaps because they have lower valuations of time relative to money, and so can afford to stand in line). We focus on proportional and efficient rationing.

We begin with the monopoly setting to illustrate how a distribution of outputs at various prices can improve upon uniform pricing and mimic second-degree price discrimination. We then explore the perfectly competitive setting to understand the effect of competition and the zero-profit condition on equilibrium pricing, as in Prescott (1975). Following Dana (1999b), we encapsulate these models into one general model of oligopoly, with monopoly and perfect competition as extremes.

8.1. Monopoly pricing with demand uncertainty and price rigidities

Consider a monopoly setting in which there are two states of aggregate demand: $D_s(q)$, $s = 1, 2$, where $D_2(q) \geq D_1(q)$. Within each demand state, consumers have unit demands but possibly different reservation prices. There is a constant marginal cost of production, c , and a cost of capacity, $k \geq 0$. The marginal cost, c , is only expended if production takes place and capacity exists; k must be spent for each unit of capacity, regardless of production. We assume that $D_1(0) > c + k$, so that the highest demanding customers value the object above the marginal production and capacity costs in the lowest demand state.

Note that in a standard non-linear pricing problem, a monopolist facing two consumer types with demands $D_1(q)$ and $D_2(q)$ would typically offer a menu of two price–quantity pairs to screen the buyers and extract more information rents on the margin. Similarly, in our aggregate demand uncertainty setting with non-contingent pricing, a monopolist would want to implement the same allocation when the consumers are replaced with mathematically equivalent markets. This is not possible, however, because of consumer heterogeneity within each demand state.⁸⁸ The monopolist, however, can still achieve a similar albeit less profitable result offering distributions of output and prices.

For simplicity, first suppose that rationing is efficient. In our two-state example, the monopolist chooses buckets of output and corresponding prices to solve

$$\max_{\{q_1, q_2\}} f_1(D_1(q_1) - c)q_1 + f_2(D_2(q_1 + q_2) - c)q_2 - k(q_1 + q_2),$$

⁸⁷ “Efficient” rationing is so named because it guarantees that there is never an ex post misallocation of goods among consumers. The term is a slight misnomer, as we will see, in that the efficient rationing rule does not guarantee that the allocation is ex post efficient between consumers and firms; that is, in most demand states some capacity will be unsold, although the marginal consumer values the output above marginal cost.

⁸⁸ The fact that the firm cannot discriminate over consumer heterogeneity is related to the discussion in Section 5.

subject to $p_1 = D_1(q_1) \leq p_2 = D_2(q_1 + q_2)$ (i.e., prices are increasing with demand). The first-order conditions (ignoring the monotonicity constraint) are

$$D_1(q_1) \left(1 - \frac{1}{\varepsilon_1(q_1)} \right) = c + k - f_2 D_2'(q_1 + q_2) q_2,$$

$$D_2(q_1 + q_2) \left(1 - \frac{1}{\varepsilon_2(q_2)} \right) = c + \frac{k}{f_2}.$$

Note that in the high-demand state, $s = 2$, the marginal cost of capacity is $\frac{k}{f_2}$; i.e., the marginal cost of capital in the high-demand state multiplied by the probability of the high-demand state must equal the cost of producing the capacity. Interestingly, the marginal price in the high-demand state is set at the state-contingent optimal monopoly price, while the price in the low-demand state is biased upward from the optimal price (and hence output is distorted downward, relative to a state-contingent monopolist). Hence, the outcome is reminiscent of second-degree price discrimination, where the monopolist distorts quality downward for the low-type consumer. Also note that if k is sufficiently large, the requirement that $p_2 \geq p_1$ is satisfied in the relaxed program; otherwise, no price dispersion arises.

Consider the case of multiplicative demand uncertainty explored in Dana (1999b) in which demand is characterized by $q = X_s(p) = sX(p)$. Holding price fixed, the elasticity of demand is not affected by demand shocks. To take a numerical example, suppose that $X(p) = 4 - p$, $s_1 = 1$, $s_2 = 2$, $c = 1$, $k = 1$; if both states are equally likely, then the profit-maximizing state-contingent prices are $p_1 = 3$ and $p_2 = \frac{7}{2}$. However, when prices cannot directly depend upon the demand state and rationing is efficient, then optimal monopoly prices are $p_1 = \frac{46}{15}$ and $p_2 = \frac{56}{15}$.⁸⁹

Now consider proportional rationing under monopoly, as in Dana (2001a). Because all consumers have an equal chance of purchasing at the low price of p_1 in the high demand state, the residual demand at p_2 is $X(p_2)(1 - \frac{q_1}{X_2(p_1)})$ at p_2 . An additional economic effect arises. It is now possible that the low-demand customers obtain some of the low-priced goods, thereby increasing the residual demand in the high-demand state. As a result, the monopolist can do better with proportional rationing than with efficient rationing.⁹⁰ The monopolist's program is to maximize

$$\max_{\{p_1, p_2\}} (p_1 - c - k)X_1(p_1) + f_2 \left(p_2 - c - \frac{k}{f_2} \right) X_2(p_2) \left(1 - \frac{X_1(p_1)}{X_2(p_1)} \right),$$

⁸⁹ Note that care must be taken in comparing these prices. With contingent pricing, all output sold in state 2 transacts at p_2 . With non-contingent pricing, the highest price of output sold in state 2 is p_2 , but some output is sold at p_1 as well.

⁹⁰ Better still, the monopolist extracts even more of the intrastate rents in the case of inefficient rationing. With one state, for example, the monopolist will extract all consumer surplus with inefficient rationing by posting one unit for each price on the demand curve.

subject to $p_2 \geq p_1$. Here, we switch to maximizing prices in order to rotate the high-state demand curve inward under proportional rationing; the switch is notationally less cumbersome. Quantities at each resulting price are determined recursively: $q_1 = X_1(p_1)$ and $q_2 = X_2(p_2)(1 - \frac{q_1}{X_2(p_1)})$.

The result that monopoly profits are higher under proportional rationing compared to efficient rationing is general. With proportional rationing, some low-demand consumers purchase the low-priced items in the high-demand state, raising the average price to high-demand customers. Returning to our previous example, the optimal monopoly prices under proportional rationing are $p_1 = 3$ and $p_2 = \frac{7}{2}$.⁹¹

8.2. Competition with demand uncertainty and price rigidities

Now consider the polar extreme of no market power – perfect competition. Assume there are S demand states, $s = 1, \dots, S$, ordered by increasing demand, with probability f_s and cumulative distribution $F(s)$. In a free-entry, perfectly competitive equilibrium, no firm can make positive expected profit for any output sold with positive probability. As Prescott (1975), Eden (1990), and Dana (1998, 1999b) have shown, this no-profit condition completely determines equilibrium prices.⁹²

Given an equilibrium distribution of output, suppose that the units offered at the price p sell in states s' and higher; it follows that the probability these units sell equals $1 - F(s' - 1)$. The zero-profit condition in this case is that $(p - c)[1 - F(s' - 1)] = k$. Hence, we can index the price by state and obtain

$$p(s) = c + \frac{k}{1 - F(s - 1)},$$

where $p(1) = c + k$. Competitive prices equal the marginal cost of production plus the marginal expected cost of capacity.

Two remarks on the perfectly competitive case are significant. First, consider the dispersion of the competitive prices. In our two-state, multiplicative-uncertainty numerical example, the competitive prices are $p_1 = c + k = 2$ and $p_2 = c + \frac{k}{1 - F(s_1)} = c + \frac{k}{f_2} = 3$. These prices are more dispersed than the monopoly prices under proportional rationing (recall, $p_1 = 3$ and $p_2 = \frac{7}{2}$). Dana (1999b) shows this result is general: competition generates more price dispersion than monopoly pricing when aggregate demand uncertainty is coupled with non-contingent pricing.

Second, except in the lowest demand state, price exceeds marginal cost, c . Because prices are inflexible ex post, this means that some firms will have unsold capacity priced above marginal cost, and consumption will be inefficiently low. Hence, regardless of

⁹¹ Again care must be taken in comparing these prices to the state-contingent prices (which are identical), as the former represent the highest purchase price in each state while the latter represent the price of all output sold in that state.

⁹² This analysis displays striking similarities to the Butters model of advertising discussed in the chapter by Bagwell (2007) in this volume.

the rationing rule, price inflexibility implies that consumption will be inefficiently low relative to the flexible-price benchmark in almost all states.

So far, we have said nothing about the rationing rule under perfect competition; the zero-profit condition is enough to determine the spot prices across states: $\{p(1), \dots, p(S)\}$.⁹³ Regardless of rationing rule, prices in a perfectly competitive market with demand uncertainty and ex post price rigidities vary by the state of aggregate demand and firms make zero profits. Prices are at *effective marginal cost*, where effective marginal cost includes the expected cost of adding capacity for a given state in addition to the familiar marginal cost of production. Of course, while capacity is priced efficiently, the market allocation is not Pareto efficient because ex post rationing may generate misallocations across consumers.

Finally, we turn to oligopoly. Dana (1999b) constructs an oligopoly model with symmetric firms in environments of multiplicative demand uncertainty and proportional rationing, which includes perfect competition and monopoly as special cases. When uncertainty is multiplicative and rationing is proportional, the residual demand function for a given firm i can be calculated given the output distributions offered by the remaining firms, $q_{-i}(p)$. Using this residual demand function, Dana (1999b) calculates the symmetric equilibrium and shows that it ranges from the monopoly setting ($n = 1$) to the perfectly competitive setting as $n \rightarrow \infty$. Remarkably, Dana (1999b) also shows that the support of prices increases as n increases, as suggested in our two-state example with perfect competition. The resulting relationship between price dispersion and competition generalizes to models of oligopoly.

The result is consistent with Borenstein and Rose's (1994) finding in the airline industry of increased price dispersion as the number of competing firms on a given route rises. Hence, the price dispersion in airline pricing may not be attributable to standard second- or third-degree price discrimination arguments, but instead represents an optimal response to aggregate demand uncertainty with price rigidities.

⁹³ A competitive equilibrium also requires that the output supplied at each price level is such that the residual demand is zero across all states. With efficient rationing, the residual demand at a price $p(s)$, given cumulative output purchased at lower prices, $Q(p(s-1))$, is $X_s(p(s)) - Q(p(s-1))$. Hence, equilibrium requires that

$$q(p(s)) = \max\{X_s(p(s)) - Q(p(s-1)), 0\},$$

with $q(p(1)) = X_1(c+k)$. Alternatively, in a world of proportional rationing, residual demand at $p(s)$ is

$$RD_s(p(s)) = X_s(p(s)) \left(1 - \sum_{j=1}^{s-1} \frac{q(j)}{X_j(p(j))} \right).$$

Equilibrium requires this to be zero in each state, which provides a recursive relationship to determine the quantities offered at each spot price.

9. Summary

While the extremes of perfect competition and monopoly are invaluable tools for understanding the economic world around us, most economic activity takes place in the realm in between these poles. One need not search very far within this sphere of imperfect competition to find numerous examples of the price discrimination strategies described in this chapter. Given the significance of these practices, an understanding of the interaction of price discrimination and competition – and how this interaction affects profits, consumer surplus, market structure and welfare – is an integral topic in industrial organization. This chapter documents many theories where (under imperfect competition) price discrimination increases welfare, providing that markets are not foreclosed. That said, even this finding is not without exceptions and counterexamples. At times, it may be frustrating that truly robust theoretical predictions are a rarity due to the additional effects of imperfect competition on our classic price-discrimination theories. In many circumstances the theories cannot provide definitive answers without additional empirical evidence. Conclusions regarding profit and welfare typically depend upon the form of consumer heterogeneity, the goods for sale, and the available instruments of price discrimination. Nonetheless, in the end the theories inform by making these dependencies clear.

The theoretical research to date also makes clear that we should not expect that the predictions of monopoly price discrimination theory will survive empirical tests using data from imperfectly competitive markets. The most interesting empirical question is that which comes after data rejects the monopoly discrimination theory: “What is the best alternative theory of price discrimination under imperfect competition?” Here, provocative combinations of theoretical and empirical work lie before us.

Acknowledgements

I am grateful to Jim Dana, Wouter Dessein, J-P. Dube, Glenn Ellison, David Genesove, Joyce Van Grondelle, David Martimort, Canice Prendergast, Markus Reisinger, Jean-Charles Rochet, Klaus Schmidt, Dan Spulber, Achim Wambach, Michael Whinston, and seminar participants at the Center for Economic Studies, Ludwig-Maximilians-Universität, Munich, Germany for helpful comments. I am especially indebted to Mark Armstrong and Audris Wong for their many helpful comments and insights. Financial support received from the National Science Foundation, the University of Chicago, GSB, and from the CES.

References

- Acquisti, A., Varian, H. (2005). “Conditioning prices on purchase history”. *Marketing Science* 24, 367–381.
Adams, W.J., Yellen, J.L. (1976). “Commodity bundling and the burden of monopoly”. *Quarterly Journal of Economics* 90, 475–498.

- Aghion, P., Bolton, P. (1987). "Contracts as a barrier to entry". *American Economic Review* 77, 388–401.
- Aguirre, I., Espinosa, M., Macho-Stadler, I. (1998). "Strategic entry deterrence through spatial price discrimination". *Regional Science and Urban Economics* 28, 297–314.
- Anderson, S., de Palma, A., Thisse, J.-F. (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge, MA.
- Anderson, S., Leruth, L. (1993). "Why firms may prefer not to price discriminate via mixed bundling". *International Journal of Industrial Organization* 11, 49–61.
- Armstrong, M. (1999). "Price discrimination by a many-product firm". *Review of Economic Studies* 66, 151–168.
- Armstrong, M. (2006). "Recent developments in the economics of price discrimination". In: Blundell, R., Newey, W., Persson, T. (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, vol. 2.
- Armstrong, M., Rochet, J.-C. (1999). "Multi-dimensional screening: A user's guide". *European Economic Review* 43, 959–979.
- Armstrong, M., Sappington, D.E.M. (2007). "Recent developments in the theory of regulation". In: Armstrong, M., Porter, R. (Eds.), *Handbook of Industrial Organization*, vol. 3. North-Holland, Amsterdam (this volume).
- Armstrong, M., Vickers, J. (1993). "Price discrimination, competition and regulation". *Journal of Industrial Economics* 41, 335–359.
- Armstrong, M., Vickers, J. (2001). "Competitive price discrimination". *RAND Journal of Economics* 32, 579–605.
- Bagwell, K. (2007). "The economic analysis of advertising". In: Armstrong, M., Porter, R. (Eds.), *Handbook of Industrial Organization*, vol. 3. North-Holland, Amsterdam (this volume).
- Baker, J. (2003). "Competitive price discrimination: The exercise of market power without anticompetitive effects (comment on Klein and Wiley)". *Antitrust Law Journal* 70, 643–654.
- Bakos, Y., Brynjolfsson, E. (1999). "Bundling information goods: Pricing, profits and efficiency". *Management Science* 45, 1613–1630.
- Bakos, Y., Brynjolfsson, E. (2000). "Bundling and competition on the Internet: Aggregation strategies for information goods". *Marketing Science* 19, 63–82.
- Banerjee, A., Summers, L. (1987). "On frequent flyer programs and other loyalty-inducing arrangements". Working Paper, No. 1337. Harvard Institute for Economic Research (HIER).
- Baumol, W., Swanson, D. (2003). "The New Economy and ubiquitous competitive price discrimination: Identifying defensible criteria of market power". *Antitrust Law Journal* 70, 661–686.
- Bernheim, D., Whinston, M. (1986). "Common agency". *Econometrica* 54, 923–943.
- Besanko, D., Dube, J.-P., Gupta, S. (2003). "Competitive price discrimination strategies in a vertical channel using aggregate retail data". *Management Science* 49, 1121–1138.
- Bester, H., Petrakis, E. (1996). "Coupons and oligopolistic price discrimination". *International Journal of Industrial Organization* 14, 227–242.
- Bhaskar, V., To, T. (2004). "Is perfect price discrimination really efficient? An analysis of free entry". *RAND Journal of Economics* 35, 762–776.
- Biais, B., Martimort, D., Rochet, J.-C. (2000). "Competing mechanisms in a common value environment". *Econometrica* 68, 799–838.
- Biglaiser, G., Mezzetti, C. (1993). "Principals competing for an agent in the presence of adverse selection and moral hazard". *Journal of Economic Theory* 61, 302–330.
- Bliss, C. (1988). "A theory of retail pricing". *Journal of Industrial Economics* 36, 375–391.
- Bond, E., Gresik, T. (1996). "Regulation of multinational firms with two active governments: A common agency approach". *Journal of Public Economics* 59, 33–53.
- Borenstein, S. (1985). "Price discrimination in free-entry markets". *RAND Journal of Economics* 16, 380–397.
- Borenstein, S., Rose, N. (1994). "Competition and price dispersion in the U.S. airline industry". *Journal of Political Economy* 102, 653–683.

- Brander, J., Eaton, J. (1984). "Product-line rivalry". *American Economic Review* 74, 323–334.
- Busse, M., Rysman, M. (2005). "Competition and price discrimination in Yellow Pages advertising". *RAND Journal of Economics* 36, 378–390.
- Cabolis, C., Clerides, S., Ioannou, I., Senft, D. (2005). "A textbook example of international price discrimination". Working Paper, July 2005. Department of Economics, University of Cyprus.
- Calem, P., Spulber, D. (1984). "Multiproduct two-part tariffs". *International Journal of Industrial Organization* 2, 105–115.
- Calzolari, G. (2004). "Incentive regulation of multinational enterprises". *International Economic Review* 45, 257–282.
- Caminal, R., Matutes, C. (1990). "Endogenous switching costs in a duopoly model". *International Journal of Industrial Organization* 8, 353–373.
- Carbajo, J., de Meza, D., Seidmann, D. (1990). "A strategic motivation for commodity bundling". *Journal of Industrial Economics* 38, 283–298.
- Carlton, D. (1978). "Market behavior with demand uncertainty and price inflexibility". *American Economic Review* 68, 571–588.
- Carlton, D., Waldman, M. (2002). "The strategic use of tying to preserve and create market power in evolving industries". *RAND Journal of Economics* 33, 194–220.
- Champsaur, P., Rochet, J.-C. (1989). "Multiproduct duopolists". *Econometrica* 57, 533–557.
- Chen, Y. (1997a). "Equilibrium product bundling". *Journal of Business* 70, 85–103.
- Chen, Y. (1997b). "Paying customers to switch". *Journal of Economics and Management Strategy* 6, 877–897.
- Chen, Y. (1999). "Oligopoly price discrimination and resale price maintenance". *RAND Journal of Economics* 30, 441–455.
- Cheung, F., Wang, X. (1999). "A note on the effect of price discrimination on entry". *Journal of Economics and Business* 51, 67–72.
- Choi, J.P., Stefanadis, C. (2001). "Tying, investment, and the dynamic leverage theory". *RAND Journal of Economics* 32, 52–71.
- Clerides, S. (2002). "Price discrimination with differentiated products: Definition and identification". *International Journal of Industrial Organization* 20, 1385–1408.
- Clerides, S. (2004). "Book value: Inter-temporal pricing and quality discrimination in the U.S. markets for books". *Economic Inquiry* 42, 402–412.
- Cohen, A. (2001). "Package size and price discrimination: Evidence from paper towels". Working Paper. University of Virginia.
- Cooper, J., Froeb, L., O'Brien, D., Tschantz, S. (2005). "Does price discrimination intensify competition? Implications for antitrust". *Antitrust Law Journal* 72, 327–373.
- Corts, K. (1998). "Third-degree price discrimination in oligopoly: All-out competition and strategic commitment". *RAND Journal of Economics* 29, 306–323.
- Cournot, A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses*. English ed. Bacon, N. (Ed.), *Researches into the Mathematical Principles of the Theory of Wealth*, MacMillan, New York, 1897.
- Courty, P., Li, H. (2000). "Sequential screening". *Review of Economics Studies* 67, 697–717.
- Crawford, G., Shum, M. (2001). "Empirical modeling of endogenous quality choice: The case of cable television". Working Paper.
- Dana Jr., J.D. (1998). "Advance-purchase discounts and price discrimination in competitive markets". *Journal of Political Economy* 106, 395–422.
- Dana Jr., J.D. (1999a). "Using yield management to shift demand when the peak time is unknown". *RAND Journal of Economics* 30, 456–474.
- Dana Jr., J.D. (1999b). "Equilibrium price dispersion under demand uncertainty: The roles of costly capacity and market structure". *RAND Journal of Economics* 30, 632–660.
- Dana Jr., J.D. (2001a). "Monopoly price dispersion under demand uncertainty". *International Economic Review* 42, 649–670.
- Dana Jr., J.D. (2001b). "Competition in price and availability when availability is unobservable". *RAND Journal of Economics* 32, 497–513.

- De Fraja, G. (1996). "Product-line competition in vertically differentiated markets". *International Journal of Industrial Organization* 14, 389–414.
- DeGraba, P. (1990). "Input market price discrimination and the choice of technology". *American Economic Review* 80, 1246–1253.
- Deneckere, R., Peck, J. (1995). "Competition over price and service rate when demand is stochastic: A strategic analysis". *RAND Journal of Economics* 26, 148–162.
- Dessein, W. (2003). "Network competition in nonlinear pricing". *RAND Journal of Economics* 34, 593–611.
- Diamond, P. (1971). "A model of price adjustment". *Journal of Economic Theory* 3, 156–168.
- Diamond, P., Maskin, E. (1979). "An equilibrium analysis of search and breach of contract. I. Steady states". *Bell Journal of Economics* 10, 282–316.
- Dupuit, J., (1849). *Annales des ponts et chaussées*, 17. Translated in: "On Tolls and Transport Charges". *International Economic Papers*, Macmillan, London, 1952.
- Economides, N. (1989). "Desirability of compatibility in the absence of network externalities". *American Economic Review* 79, 1165–1181.
- Eden, B. (1990). "Marginal cost pricing when spot markets are complete". *Journal of Political Economy* 98, 1293–1306.
- Ellison, G. (2005). "A model of add-on pricing". *Quarterly Journal of Economics* 120, 585–637.
- Farrell, J., Klemperer, P. (2007). "Coordination and lock-in: Competition with switching costs and network effects". In: Armstrong, M., Porter, R. (Eds.), *Handbook of Industrial Organization*, vol. 3. North-Holland, Amsterdam (this volume).
- Fudenberg, D., Tirole, J. (1991). *Game Theory*. MIT Press, Cambridge, MA.
- Fudenberg, D., Tirole, J. (2000). "Customer poaching and brand switching". *RAND Journal of Economics* 31, 634–657.
- Fudenberg, D., Villas-Boas, M. (2005). "Behavior-based price discrimination and customer recognition". In: *Handbook of Economics and Information Systems*. North-Holland, Amsterdam. In press.
- Gabszewicz, J., Thisse, J.-F. (1979). "Price competition, quality and income disparities". *Journal of Economic Theory* 20, 340–359.
- Gabszewicz, J., Thisse, J.-F. (1980). "Entry, (and exit) in a differentiated industry". *Journal of Economic Theory* 22, 327–338.
- Gal-Or, E. (1983). "Quality and quantity competition". *The Bell Journal of Economics* 14 (2), 590–600.
- Gal-Or, E. (1991). "A common agency with incomplete information". *RAND Journal of Economics* 22, 274–286.
- Gale, I., Holmes, T. (1992). "The efficiency of advance-purchase discounts in the presence of aggregate demand uncertainty". *International Journal of Industrial Organization* 10, 413–437.
- Gale, I., Holmes, T. (1993). "Advance-purchase discounts and monopoly allocation of capacity". *American Economic Review* 83, 135–146.
- Galera, F. (2003). "Welfare and output in third-degree price discrimination: A note". Working Paper.
- Gilbert, R., Matutes, C. (1993). "Product-line rivalry with brand differentiation". *Journal of Industrial Economics* 41, 223–240.
- Goldberg, P. (1995). "Product differentiation and oligopoly in international markets: The case of the U.S. automobile industry". *Econometrica* 63, 891–951.
- Goldberg, P., Verboven, F. (2005). "Market integration and convergence to the law of one price: Evidence from the European car market". *Journal of International Economics* 65, 49–73.
- Graddy, K. (1995). "Testing for imperfect competition at the Fulton fish market". *RAND Journal of Economics* 26, 75–92.
- Gul, F. (1987). "Noncooperative collusion in durable goods oligopoly". *RAND Journal of Economics* 18, 248–254.
- Holmes, T. (1989). "The effects of third-degree price discrimination in oligopoly". *American Economic Review* 79, 244–250.
- Holton, R. (1957). "Price discrimination at retail: The supermarket case". *Journal of Industrial Economics* 6, 13–31.

- Hotelling, H. (1929). "Stability in competition". *Economic Journal* 39, 41–57.
- Hurdle, G., McFarland, H. (2003). "Criteria for identifying market power: A comment on Baumol and Swanson". *Antitrust Law Journal* 70, 687–696.
- Ivaldi, M., Martimort, D. (1994). "Competition under nonlinear pricing". *Annales d'Economie et de Statistique* 34, 71–114.
- Johnson, J.P., Myatt, D.P. (2003). "Multiproduct quality competition: Fighting brands and product-line pruning". *American Economic Review* 93, 748–774.
- Jullien, B. (2000). "Participation constraints in adverse selection models". *Journal of Economic Theory* 93, 1–47.
- Katz, M. (1984). "Price discrimination and monopolistic competition". *Econometrica* 52, 1453–1471.
- Katz, M. (1985). "Firm-specific differentiation and competition among multiproduct firms". *Journal of Business* 57, S149–S166.
- Katz, M. (1987). "The welfare effects of third-degree price discrimination in intermediate good markets". *American Economic Review* 77, 154–167.
- Klein, B., Wiley, J. (2003a). "Competitive price discrimination as an antitrust justification for intellectual property refusals to deal". *Antitrust Law Journal* 70, 599–642.
- Klein, B., Wiley, J. (2003b). "Market power in economics and in antitrust: Reply to Baker". *Antitrust Law Journal* 70, 655–660.
- Klemperer, P. (1992). "Equilibrium product lines: Competing head-to-head may be less competitive". *American Economic Review* 82, 740–755.
- Klemperer, P., Meyer, M. (1986). "Price competition vs. quantity competition: The role of uncertainty". *RAND Journal of Economics* 17, 618–638.
- Klemperer, P., Meyer, M. (1989). "Supply-function equilibria in oligopoly under uncertainty". *Econometrica* 57, 1243–1278.
- Kotler, P. (1994). *Marketing Management*, eighth ed. Prentice-Hall, New York.
- Laffont, J.-J., Tirole, J. (1991). "Privatization and incentives". *Journal of Law, Economics, and Organization* (Special Issue) 7, 84–105.
- Laffont, J.-J., Rey, P., Tirole, J. (1998). "Network competition: Price discrimination". *RAND Journal of Economics* 29, 38–56.
- Lal, R., Matutes, C. (1994). "Retail pricing and advertising strategies". *Journal of Business* 67, 345–370.
- Lederer, P., Hurter, A. (1986). "Competition of firms: Discriminatory pricing and location". *Econometrica* 54, 623–640.
- Leontief, W. (1940). "The theory of limited and unlimited discrimination". *Quarterly Journal of Economics* 54, 490–501.
- Leslie, P. (2004). "Price discrimination in Broadway Theatre". *RAND Journal of Economics* 35, 520–541.
- Lewis, T., Sappington, D. (1989). "Countervailing incentives in agency problems". *Journal of Economic Theory* 49, 294–313.
- Liu, Q., Serfes, K. (2004). "Quality of information and oligopolistic price discrimination". *Journal of Economics and Management Strategy* 13, 671–702.
- Locay, L., Rodriguez, A. (1992). "Price discrimination in competitive markets". *Journal of Political Economy* 100, 954–965.
- Lott, J., Roberts, R. (1991). "A guide to the pitfalls of identifying price discrimination". *Economic Inquiry* 29, 14–23.
- MacLeod, B., Norman, G., Thisse, J.-F. (1988). "Price discrimination and equilibrium in monopolistic competition". *International Journal of Industrial Organization* 6, 429–446.
- Maggi, G., Rodriguez-Clare, A. (1995). "On countervailing incentives". *Journal of Economic Theory* 66, 238–263.
- Mankiw, N.G., Whinston, M. (1986). "Free entry and social inefficiency". *RAND Journal of Economics* 17, 48–58.
- Martimort, D. (1992). "Multi-principaux avec anti-selection". *Annales d'Economie et de Statistique* 28, 1–38.
- Martimort, D. (1996). "Exclusive dealing, common agency, and multi-principals incentive theory". *RAND Journal of Economics* 27, 1–31.

- Martimort, D., Stole, L. (2003). "Contractual externalities and common-agency equilibria". In: *Advances in Theoretical Economics*, vol. 3. Berkeley Electronic Press. <http://www.bepress.com>. Article 4.
- Martimort, D., Stole, L. (2005). "Market participation under delegated and intrinsic common agency games". Working Paper.
- Maskin, E., Riley, J. (1984). "Monopoly with incomplete information". *RAND Journal of Economics* 15, 171–196.
- Matutes, C., Regibeau, P. (1988). "Mix and match: Product compatibility without network externalities". *RAND Journal of Economics* 19, 221–234.
- Matutes, C., Regibeau, P. (1992). "Compatibility and bundling of complementary goods in a duopoly". *Journal of Industrial Economics* 40, 37–54.
- McAfee, P., McMillan, J., Whinston, M. (1989). "Multiproduct monopoly, commodity bundling, and correlation of values". *Quarterly Journal of Economics* 104, 371–383.
- McManus, B. (2004). "Nonlinear pricing in an oligopoly market: The case of specialty coffee". Working Paper. Olin School of Business, Washington University.
- Mezzetti, C. (1997). "Common agency with horizontally differentiated principals". *RAND Journal of Economics* 28, 323–345.
- Miravete, E. (1996). "Screening consumers through alternative pricing mechanisms". *Journal of Regulatory Economics* 9, 111–132.
- Miravete, E., Röller, L.-H. (2003). "Competitive nonlinear pricing in duopoly equilibrium: The early U.S. cellular telephone industry". Working Paper.
- Mussa, M., Rosen, S. (1978). "Monopoly and product quality". *Journal of Economic Theory* 18, 301–317.
- Nahata, B., Ostaszewski, K., Sahoo, P. (1990). "Direction of price changes in third-degree price discrimination". *American Economic Review* 80, 1254–1258.
- Nalebuff, B. (2004). "Bundling as an entry barrier". *Quarterly Journal of Economics* 119, 159–187.
- Nevo, A., Wolfram, C. (2002). "Why do manufacturers issue coupons?: An empirical analysis of breakfast cereals". *RAND Journal of Economics* 33, 319–339.
- Nilssen, T. (1992). "Two kinds of consumer switching costs". *RAND Journal of Economics* 23, 579–589.
- Norman, G. (1981). "Spatial competition and spatial price discrimination". *Review of Economic Studies* 48, 97–111.
- Oren, S., Smith, S., Wilson, R. (1983). "Competitive nonlinear tariffs". *Journal of Economic Theory* 29, 49–71.
- Phlips, L. (1983). *The Economics of Price Discrimination*. Cambridge Univ. Press, Cambridge, UK.
- Pigou, A.C. (1920). *The Economics of Welfare*. Macmillan, London.
- Prescott, E. (1975). "Efficiency of the natural rate". *Journal of Political Economy* 83, 1229–1236.
- Reisinger, M. (2003). "The effects of product bundling in duopoly". Working Paper.
- Rey, P., Tirole, J. (2007). "A primer on foreclosure". In: Armstrong, M., Porter, R. (Eds.), *Handbook of Industrial Organization*, vol. 3. North-Holland, Amsterdam (this volume).
- Robinson, J. (1933). *The Economics of Imperfect Competition*. Macmillan, London.
- Rochet, J.-C., Stole, L. (2002). "Nonlinear pricing with random participation". *Review of Economic Studies* 69, 277–311.
- Rochet, J.-C., Stole, L. (2003). "The economics of multidimensional screening". In: Dewatripont, M., Hansen, L. (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, vol. 1. Cambridge Univ. Press.
- Rosenthal, R. (1980). "A model in which an increase in the number of sellers leads to a higher price". *Econometrica* 48, 1575–1580.
- Rossi, P., Allenby, G. (1993). "A Bayesian approach to estimating household parameters". *Journal of Marketing Research* 30, 171–182.
- Rossi, P., McCulloch, R., Allenby, G. (1996). "The value of purchase history data in target marketing". *Marketing Science* 15, 321–340.
- Salop, S. (1977). "The noisy monopolist: Imperfect information, price dispersion and price discrimination". *Review of Economic Studies* 44, 393–406.

- Salop, S. (1979). "Monopolistic competition with outside goods". *Bell Journal of Economics* 10, 141–156.
- Salop, S., Stiglitz, J. (1977). "Bargains and ripoffs: A model of monopolistically competitive price dispersion". *Review of Economic Studies* 44, 493–510.
- Salop, S., Stiglitz, J. (1982). "A theory of sales: A simple model of equilibrium price dispersion with identical agents". *American Economic Review* 72 (5), 1121–1130.
- Schmalensee, R. (1981). "Output and welfare effects of monopolistic third-degree price discrimination". *American Economic Review* 71, 242–247.
- Schmalensee, R. (1984). "Gaussian demand and commodity bundling". *Journal of Business* 57, S211–S230.
- Schmidt-Mohr, U., Villas-Boas, M. (1999). "Oligopoly with asymmetric information: Differentiation in credit markets". *RAND Journal of Economics* 30, 375–396.
- Schwartz, M. (1990). "Third-degree price discrimination and output: Generalizing a welfare result". *American Economic Review* 80, 1259–1262.
- Shaffer, G., Zhang, Z.J. (1995). "Competitive coupon targeting". *Marketing Science* 14, 395–416.
- Shaffer, G., Zhang, Z.J. (2000). "Pay to switch or pay to stay: Preference-based price discrimination in markets with switching costs". *Journal of Economics & Management Strategy* 9, 397–424.
- Shaked, A., Sutton, J. (1982). "Relaxing price competition through product differentiation". *Review of Economic Studies* 49 (1), 3–13.
- Shaked, A., Sutton, J. (1983). "Natural oligopolies". *Econometrica* 51 (5), 1469–1484.
- Shepard, A. (1991). "Price discrimination and retail configuration". *Journal of Political Economy* 99, 30–53.
- Spence, M. (1976a). "Product differentiation and welfare". *Papers and Proceedings of the American Economic Review* 66, 407–414.
- Spence, M. (1976b). "Product selection, fixed costs, and monopolistic competition". *Review of Economic Studies* 43, 217–235.
- Spulber, D. (1979). "Noncooperative equilibrium with price discriminating firms". *Economic Letters* 4, 221–227.
- Spulber, D. (1984). "Competition and multiplant monopoly with spatial nonlinear pricing". *International Economic Review* 25, 425–439.
- Spulber, D. (1989). "Product variety and competitive discounts". *Journal of Economic Theory* 48, 510–525.
- Stahl, D. (1989). "Oligopolistic pricing with sequential consumer search". *American Economic Review* 79, 700–712.
- Stigler, G. (1963). "United States v. Loew's Inc.: A note on block-booking". *The Supreme Court Review* 1, 152–157.
- Stigler, G. (1987). *A Theory of Price*. Macmillan, New York.
- Stiglitz, J. (1989). "Imperfect information in the product market". In: Schmalensee, R., Willig, R.D. (Eds.), *Handbook of Industrial Organization*, vol. 1. North-Holland, Amsterdam, pp. 769–847. Distributed in the U.S. and Canada by Elsevier Science, New York.
- Stole, L. (1991). "Mechanism design and common agency". Working Paper.
- Stole, L. (1995). "Nonlinear pricing and oligopoly". *Journal of Economics & Management Strategy* 4, 529–562.
- Taylor, C. (2003). "Supplier surfing: Competition and consumer behavior in subscription markets". *RAND Journal of Economics* 34, 223–246.
- Thisse, J.-F., Vives, X. (1988). "On the strategic choice of spatial price policy". *American Economic Review* 78, 122–137.
- Tirole, J. (1988). *The Theory of Industrial Organization*. MIT Press, Cambridge, MA.
- Varian, H. (1980). "A model of sales". *American Economic Review* 70, 651–659.
- Varian, H. (1985). "Price discrimination and social welfare". *American Economic Review* 75, 870–875.
- Varian, H. (1989). "Price discrimination". In: Schmalensee, R., Willig, R.D. (Eds.), *Handbook of Industrial Organization*, vol. 1. North-Holland, Amsterdam, pp. 597–654. Distributed in the U.S. and Canada by Elsevier Science, New York.
- Verboven, F. (1996). "International price discrimination in the European car market". *RAND Journal of Economics* 27, 240–268.

- Verboven, F. (1999). "Product-line rivalry and market segmentation – With an application to automobile optional engine pricing". *Journal of Industrial Economics* 47, 399–426.
- Villas-Boas, M. (1999). "Dynamic competition with customer recognition". *RAND Journal of Economics* 30, 604–631.
- Ward, M. (2003). "Symposium: Competitive price discrimination (Editor's note)". *Antitrust Law Journal* 70, 593–598.
- Whinston, M. (1990). "Tying, foreclosure, and exclusion". *American Economic Review* 80, 837–859.
- Wilson, R. (1993). *Nonlinear Pricing*. Oxford Univ. Press, Oxford, UK.
- Winter, R. (1997). "Colluding on relative prices". *RAND Journal of Economics* 28, 359–371.
- Yoshida, Y. (2000). "Third-degree price discrimination in input markets: Output and welfare". *American Economic Review* 90, 240–246.