

R. Betancourt (Draft: February 2000).

## Chapter 2: The Economic Function of Retail Organizations

A main theme of this work is that the economic function of any retail organization is to provide consumers with a set of distribution services together with the explicit items or services bought at retail. In Section 2.1 we discuss six types of distribution or transaction costs incurred by consumers and how they map into five distribution services or outputs provided by retail organizations. One implication of this discussion is that the shifting of distribution or transaction costs between retailers and consumers is an essential characteristic of retail markets. This characteristic is discussed in Section 2.2, where we also explain the two formal economic concepts that capture cost shifting: namely, distribution services as outputs of retail institutions and as fixed inputs into the purchase or consumption activities of consumers. Subsequently, (Section 2.3) we discuss a second essential characteristic of retail markets: the bundling of these distribution services among themselves and with the items or explicit services provided at retail. We also present here basic implications of these characteristics for the specification of demand and cost functions.

Characteristics such as the above two are elementary but powerful in helping our understanding of retail markets. For instance, they are sufficient to generate price dispersion and product variety with respect to distribution services in these markets, which is illustrated with a simple model of monopolistic competition in Section 2.4. Furthermore, they have profound implications for competition and welfare in these markets. Some of these are derived in Section 2.5. by means of a model that provides the

basis for analysis throughout most of the book. In Section 2.6 we demonstrate how this model generates the so called full price model of services as a special case, which also brings out two limitations of the full price model. Finally, in Section 2.7, we discuss the measurement of distribution services and empirical evidence on their role in explaining retail margins at the sectoral level.

### 2.1. Distribution Costs and Services.<sup>1</sup>

In order to carry out their purchase or consumption activities consumers incur a variety of costs that can be characterized as distribution or transaction costs. We begin by identifying six types of distribution costs that consumers can incur when interacting with any part of the retail system. Among the most easily identifiable ones are, of course, direct time and transportation costs. The former include the opportunity costs of travel time to and from a retail establishment, waiting time inside or outside the establishment and time spent in planning purchase activities. The latter include the monetary cost of transport to and from the purchase site. These costs are obvious enough that they require no further elaboration.

Purchasing or consumption activities also generate distribution costs in the form of adjustment costs as a result of the unavailability of products or services at the desired time of consumption or purchase. These adjustment costs arise as a result of indirect time and transportation costs incurred due to forced search or due to the increased expenditures or lower utility associated with altering the consumption or purchase bundle of goods and services. Economists would also characterize these costs as the costs of rationing through

---

<sup>1</sup>This section draws heavily from our earlier work ( Betancourt and Gautschi, 1988).

unavailability. While not as obvious as the previous ones, they are ubiquitous and play an important role, for example, in the development of one stop shopping institutions and in the success of technological innovations such as ATM's.

One type of distribution costs stressed in the retailing literature ( Ingene, 1984) are psychic costs. These are costs inflicted on the consumer utilizing the retail system by undesirable characteristics of the retail environment. Examples of these characteristics would be drudgery, anxiety or disagreeable social interactions.

Another type of distribution costs are storage costs. These would arise, for instance, in the purchase of household products in bulk. An example would be purchasing wine in cases rather than in bottles. This type of cost featured prominently in the development of a particular retail form, warehouse stores. The last type of cost to be considered explicitly are information costs. Information may be desired with respect to price, availability, physical attributes or performance characteristics of the goods and services provided at the retail level as well as with respect to similar characteristics of the retail establishment. The acquisition of this information entails costs through the use of time, transport and other resources.

Having considered the demand or end use side of the market, it is useful to switch perspective and consider the supply or retail system side. Any particular configuration of a retail system imposes a particular level of these six types of distribution or transaction costs on consumers. Furthermore, different retail forms entail different levels of these costs for consumers. In order to make these points clearly, it is convenient to think in terms of five broad categories of distribution services that are provided by any retail

organization, albeit at different levels. There is some degree of arbitrariness in the level of detail at which one chooses to describe these services. Our particular choice here of five categories and the implied level of aggregation is aimed at providing a framework in which any particular type of retail establishment can be included, establishing the link between these services and the costs just described for the consumer and leaving no doubt that the production of these services requires resources. At various points in the arguments throughout the book, however, we will disaggregate some of these five services and at other points we will aggregate them into a single service.

Distribution services are difficult to define and measure with precision. Perhaps the most difficult to define and measure is what may be labeled *ambiance*. It determines the level of psychic costs imposed on the consumer by the nature of the retail environment. While this service is difficult to define and measure, it is clear that some types of retail establishments specialize in providing low and others high levels of this service. Discount stores are an example of the former where the associated lower prices of the goods sold in these establishments are due, at least in part, to the resources saved by not providing *ambiance*. By contrast, high scale stores such as Nieman Marcus are examples of the latter type of establishment. The higher prices associated with the products sold there are due, at least in part, to the cost of the resources used in providing *ambiance*, including those of operating sites in high rent districts.

A second category of distribution services provided by a retail system is the level of product assortment, which for some purposes can be subdivided into breadth (different product lines) and depth (different varieties within a product line). This service affects

the levels of several types of distribution costs experienced by consumers who patronize retail institutions. For example, it affects direct time and transportation costs associated with multiple shopping trips. It also affects adjustment costs associated with foregoing an item for purchase or consumption. Examples of retail activities stressing product breadth are supermarkets, department stores and hypermarkets. Examples of retail activities stressing product depth are specialty stores of various kinds. For a given number of goods sold, higher levels of assortment will entail higher costs for the retailer, for instance the costs of labeling and layout.

Our next category of distribution services is accessibility of location. It is the easiest one to define and measure in many situations. At the most elementary level it is the distance to the retail establishment. It affects the direct time and transportation costs experienced by consumers in their purchasing or consumption activities. For a given number of goods sold, a retail system may provide greater accessibility to consumers by having several retail sites in a given market area, and this configuration would entail higher costs than operating a single retail site.

A fourth type of output provided by the retail system is assurance of product delivery, which can be disaggregated into delivery at the desired time and in the desired form. The former is accomplished through opening hours as well as through the provision of credit, for example. The latter includes functions such as breaking bulk as well as risk bearing through the acquisition of ownership or the provision of warranties. This service affects several types of distribution costs: the direct time costs of waiting inside or outside the establishment; adjustment costs due to unavailability, and storage costs forced upon

the consumer by the lack of availability at the desired time in the desired quantities. The provision of higher levels of this service leads to higher cost for the providers in the retail system in the form of costs of extended opening hours, credit provision, storage and risk bearing.

Last but not least among the distribution services provided by the retail system is the amount of information with respect to prices, availability and other characteristics of the goods and services provided as well as of the retail establishment. The provision of higher levels of this service through advertising and sales personnel, for example, leads to higher costs for the retail system. On the other hand, it lowers consumers' costs of information, adjustment and storage.

Several implications of the previous discussion of distribution costs and services are worth emphasizing at this point. First, there is no necessary one to one mapping between the six types of distribution costs and the five types of distribution services. Since we concentrate on distribution services in what follows, we will not try to provide an exhaustive account of the possible mappings here. Second, there is jointness in the provision of distribution services. For instance, by providing information on store hours a retailer increases assurance of product delivery at the desired time as well; similarly, by providing deeper assortments a retailer increases assurance of product delivery in the desired form. Finally, some distribution services are specific to a product or line of products whereas others are common to the entire assortment. For example, by having two identical stores instead of one a retailer is providing higher levels of accessibility of location to the entire assortment. Similarly by adding a new product line the retailer is

expanding the breadth of assortment for all the items that he or she carries. Hence, these two services are examples of common distribution services. By contrast, other services such as information on the nature of a specific item are of value only to the item at issue. For example, how well a suit or a shoe fits or the operating properties of an appliance are examples of information specific to the item or the product line.

## 2.2. Cost Shifting.

One implication of the previous discussion is that cost shifting is an essential characteristic of retail markets. Thus, it will be the focus of this section and it will play an important role throughout the book. In each of the five categories of distribution services identified above, we have illustrated how providing higher levels of any of these distribution services entails higher costs for the retailers. We have also illustrated how the higher levels of these services lower different types of distribution costs incurred by consumers. That is we have illustrated the shifting of costs that takes place in retail markets between retailers and consumers.

The idea of cost shifting is not novel either in the economics or in the marketing literature. For instance, in economics Fuchs (1968) attributed the productivity gains of supermarkets to putting the consumers to work; in marketing Ingene (1984) discusses productivity in terms of the shifting of functions between consumers and retailers. What the previous discussion makes clear, however, is that this is not a limited phenomenon but a pervasive one that takes place in all the possible dimensions of distribution services that a retail system provides. Hence, it can be misleading to discuss productivity by looking at only one of these dimensions.

Considering the pervasiveness of cost shifting in retail markets, it is useful to be as rigorous and precise as possible in capturing this notion. On the side of the retail system this can be made precise by identifying each of these distribution services as an output of any retail organization. The fundamental economic characteristic of any output is that providing higher levels of the output entails higher levels of cost. This criterion is clearly satisfied by each of the distribution services identified above. Once again the idea of thinking about the functions of marketing systems in terms of outputs is not new. For instance, Bucklin (1973) identifies four indexes of output for the distribution sector as a whole. These collapse to accessibility, assortment and a more limited definition of assurance of product delivery. The identification of these five categories of distribution services as outputs of the retail system is somewhat recent (Betancourt and Gautschi, 1986, 1988). Nonetheless, it can be viewed as an application and extension of Bucklin's basic idea. This approach of treating distribution services explicitly as outputs has also been adopted by Oi (1992), who creates a list that overlaps with the one above.

A decision to treat each of these services as outputs implies that any retail organization that provides a higher (lower) level of any of these services is shifting distribution costs onto (away from) itself and away from (onto) its customers. Thus, it is desirable to have an equally fundamental concept as output on the demand side that can allow us to capture the shifting of costs between consumers and retailers. This concept is that of a fixed input.

Each of these distribution services that are outputs of retail organizations will be viewed as a fixed input into the production functions of end users, especially into the

household production functions of consumers. The fundamental economic characteristic of a fixed input is that higher levels of this input lower or at least do not increase the costs of producing any given level of output.<sup>2</sup> Hence, this concept captures the description of how each of the distribution services identified in the previous section affects the various types of distribution costs experienced by consumers. Thus, these two concepts provide us with a precise mechanism to capture the shifting of costs between consumers and retailers that permeates retail markets. An increase in a distribution service by a retailer increases the retailers costs in its role as an output and it usually decreases the distribution costs of any consumer who chooses to patronize the retailer in its role as a fixed input into the ( household) production function of the consumer.<sup>3</sup>

An advantage of this conceptualization is that it brings out the role played by both participants in the retail market. The consumer chooses a level of each of these distribution services when she chooses to patronize a particular retailer. The retailer chooses the levels of each distribution service that he provides by its choice of mode of operation and the characteristics of the market will determine the equilibrium configuration that is actually observed. Another advantage of this conceptualization is that it brings out the role of a retailer as a multiproduct producer with respect to

---

<sup>2</sup>This characteristic is one of the defining properties of any restricted cost function ( Fuss and McFadden, 1978).

<sup>3</sup> In some situations a consumer may not take advantage of the fixed input provided by a retailer. For instance, a consumer may not avail herself of the information provided by a sales assistant. This is perfectly consistent with the definition of the distribution service as a fixed input as it does not directly increase the costs to the consumer of ignoring this information. All we are assuming is that there is free disposal of the distribution service by the consumer. In so far as the prices of the items sold by a retailer increase because of the availability of assistance, the costs to the consumer will be affected indirectly and one would expect this particular form of service to be eliminated if many consumers find it irrelevant. Nevertheless, this set of circumstances is logically consistent with the definition of a fixed input.

distribution services. This makes it immediately clear that statements about productivity require controlling for the level or the mix of each of these outputs in some fashion. To illustrate, saying that the increase in productivity of supermarkets is due to making the consumer work more (shifting costs to the consumers) ignores the considerable lowering of distribution costs to the consumers from the larger and deeper assortments of supermarkets as well as from their convenient parking facilities. The fact that these retail forms have not only succeeded but seem to be expanding along these dimensions would suggest that the net result is a lowering of the overall distribution costs experienced by the consumer.

### 2.3. Bundling.

All five of the distribution services discussed above are bundled together in any retail setting, which generates one type of bundling. A second type of bundling, however, is generated by the bundling of these distribution services with an additional indispensable output of the retail system left out from the previous discussion -- the explicit goods or services distributed by the retailer. This type of bundling generates another essential feature of retail markets: namely, the consumer explicitly pays for these explicit items or services and only implicitly for the outputs of distribution services that are bundled with them.<sup>4</sup> One should also note that there can be jointness in the provision of a distribution service and in the provision of explicit outputs. For instance, an increase in assortment usually entails an increase in the number of items to be distributed by the

---

<sup>4</sup>There can be exceptions, of course, as in the cases where delivery services are explicitly charged for when a customer employs them or when a gas station charges differentially for self-service and full-service.

retailer. To sum up, in viewing the retailer as a multiproduct producer we have one more dimension of output and this dimension is different from the others in that it is the one that is explicitly priced.<sup>5</sup>

For purposes of analysis this dimension of output must be treated differently from the others. There are two ways of specifying the demand for the explicit output of a retailer depending on the assumptions one makes about the retailer's decision variables. One can assume that the retailer chooses the quantities of the items for sale. In that case the demand function faced by the retailer is specified as an inverse demand function . For instance,

$$(2.1) \quad p^* = f(Q, D, W, p')$$

where  $p^*$  is the retailer's price,  $Q$  is the quantity of explicit items for sale,  $D$  is a vector of outputs representing the five distribution services identified previously,  $W$  is the full income of a representative consumer and  $p'$  is a set of other prices that can potentially affect the consumer's demand. This demand function has the following properties: It is a nonincreasing function of the quantity of items distributed ( $f_Q \leq 0$ ), a nondecreasing function of each of the distribution services provided ( $f_D \geq 0$ , for all elements of  $D$ ), and a nondecreasing function of full income ( $f_W \geq 0$ ).<sup>6</sup> Thus, consumers are willing to pay

---

<sup>5</sup>Incidentally, retailers sometimes engage in production as well as distribution of the explicit items sold. In these cases it is frequently impossible to separate the distribution activity from the production activity. This happens frequently in the case of services. For example in the case of restaurants production and distribution of a meal are usually inseparable. Hence, the explicit output of the restaurant would be the meals. The five categories of distribution services would continue to be relevant. The main practical consequence of inseparability between the explicit output of distribution and its production would be that the value added generated by a restaurant would be greater than of a typical retail form that did not engage in production activities. This would be so precisely because of the contribution of production activities to value added in the case of restaurants.

<sup>6</sup>These properties are derived in Betancourt and Gautschi (1988).

the same or a higher price for a given set of items (Q) that come in a bundle with at least one distribution service at a higher level and all the others the same than another bundle. The other two properties are standard ones for inverse demand functions.

In some instances we want to specify the retailer as choosing prices rather than quantities. In these cases the demand function faced by the retailer would be specified in standard form as

$$(2.2) Q = g(p^*, D, W, p')$$

The variables are the same as before. This demand function would have the following properties: It is a nonincreasing function of the retail price ( $g_{p^*} \leq 0$ ), a nondecreasing function of each of the distribution services provided ( $g_D \geq 0$ , for all elements of D), and a nondecreasing function of full income ( $g_W \geq 0$ ). In this case consumers are willing to buy the same or more explicit items at a given price when they come in a bundle with at least one distribution service at a higher level and all the others the same than in another bundle. The other two properties are standard ones for ordinary demand functions.<sup>7</sup>

The above two specifications capture the consequences of the bundling of explicit outputs with distribution services on the demand side. With respect to the supply side, this bundling merely implies that a retailer is a multiproduct producer and any special features of the bundling can be captured through the properties of a multiproduct cost function, which can be specified in general as follows:

$$(2.3) C = C(Q, D, v)$$

where C are the costs of retailing, v is a vector of input prices and Q and D are as defined

---

<sup>7</sup>In the next chapter we discuss in greater detail the derivation of this demand function.

previously. The main property of this function that interest us at this point is that it is increasing in outputs. That is,  $C_Q > 0$  and  $C_D > 0$ , for all elements of  $D$ .<sup>8</sup>

#### 2.4. Price Dispersion and Product Variety in Distribution Services.

In this section we illustrate two important consequences of bundling in terms of a simple model of monopolistic competition. First, the bundling of distribution services with the explicit output sold at retail is a fundamental source of equilibria with price dispersion and differential cost shifting in retail markets. Second, the bundling of distribution services among themselves is a fundamental source of equilibria with product variety in distribution services and differential cost shifting in retail markets.

Consider a representative profit-maximizing retailer who faces the demand function in (2.1), i.e., a quantity setter, and who is subject to the cost function in (2.3). Her objective is to maximize the profits ( $\pi$ ) of retailing activities. Thus,

$$(2.4) \pi = p^*Q - C(Q, D, v) - pQ,$$

where  $p$  is the price of the items acquired from suppliers. Hence,  $pQ$  is the costs of goods sold. To keep matters simple we will assume that there is only one distribution service subject to the retailer's control,  $D$ .

The first-order conditions for profit maximization can be written, after some manipulation, as

$$(2.5) (p^* - p)Q/C - S_Q = p^*\epsilon_Q/C$$

$$(2.6) S_D C/Q = p^*\epsilon_D$$

where  $S_Q$  is the proportionate increase in costs from distributing an additional item

---

<sup>8</sup>In Chapter 4 we discuss the underpinnings of the cost function and its properties in greater detail.

$(C_Q Q/C)$  and  $S_D$  is the proportionate increase in costs from providing an additional unit of the distribution service  $(C_D D /C)$ .  $\varepsilon$  is the absolute value of the price elasticity of demand with respect to the items for sale  $(-f_Q Q/p^*)$  and  $\varepsilon_D$  is the distribution services elasticity of demand for the  $i$ th distribution service  $(f_D D /p^*)$ .

If the market is to be in long-run equilibrium profits must be zero. Therefore, this condition implies from (2.4) that  $(p^*-p) = C/Q$ . In other words the retail margin obtained from selling an additional item must equal the average costs of retailing this additional item. Imposing this condition on the short-run equilibrium ones leads to

$$(2.7) \quad (p^*-p)/p^* = \varepsilon_D / S_D = \varepsilon / (1- S_Q).$$

The left hand side of this equation is the retail gross margin expressed as a percentage of sales,  $R$ , and it moves in the same direction as the retail price.

Equation (2.7) can be used to generate the following proposition:

*Proposition 1.* Heterogeneity among consumers with respect to price sensitivity or the demand for distribution services is sufficient to yield equilibria with price dispersion and differential cost shifting even if firms are identical.

To illustrate the validity of this proposition consider first two market segments, A and B, where consumers differ with respect to their price sensitivity by having different values of a constant price elasticity of demand:  $\varepsilon(A) > \varepsilon(B)$ .<sup>9</sup> We will assume the cost function for firms to be such that  $S_D = (1-S_Q) = k$ , where  $k$  is a constant.<sup>10</sup> Long run equilibrium in segments A and B requires, by (2.7), that the retail gross margins,  $R$ , be

---

<sup>9</sup>Note that we have an inverse demand function; hence  $\varepsilon(A) > \varepsilon(B)$  implies a greater price sensitivity in B than in A when measured in the usual way, i.e., in terms of the reciprocals.

<sup>10</sup>This can be generated by the following function:  $C = c(v)D^\alpha Q^\beta$ , where  $v$  are input prices and  $\alpha = (1-\beta)$ .

such that  $R(A) = [\varepsilon_D(A) = \varepsilon(A)]/k > R(B) = [\varepsilon_D(B) = \varepsilon(B)]/k$ . If the inverse demand function has the following form,  $p^* = h(D)Q^{-\varepsilon}W$ , then  $\varepsilon_D = (h_D D)/h(D)$ . Note that  $h_D = \partial h(D)/\partial D$ . The retail price will always be higher in A than in B. If the distribution services elasticity is decreasing (increasing) in D, however, the level of distribution services will be lower (higher) in A than in B. In either case there will be price dispersion and a different level of distribution costs (differential cost shifting) borne by consumers in long run equilibrium.

When consumer heterogeneity is generated by differences in the valuation of distribution services, we can establish a similar conclusion. For this purpose, let the inverse demand function have the following form,  $p^* = D^\delta m(Q)W$ . The price elasticity of demand is now variable, i.e.,  $\varepsilon = -[m_Q Q/m(Q)]$ , and the distribution services elasticity of demand is now constant, i.e.,  $\varepsilon_D = \delta$ . If in segment A consumers value distribution services more than in segment B,  $\delta(A) > \delta(B)$ . This situation can arise because the opportunity cost of their time, for example, is different.<sup>11</sup> It follows from (2.7), under the same assumptions on k as before, that  $R(A) = [\delta(A) = \varepsilon(A)]/k > R(B) = [\delta(B) = \varepsilon(B)]/k$ . Thus, the retail price will be higher in the segment with the higher valuation of distribution services. Moreover, there is a wide array of combinations of distribution services, D, and output, Q, that are consistent with this equilibrium, including many which entail differential cost shifting in the two markets.

Consumer heterogeneity as a source of equilibrium price dispersion in retail

---

<sup>11</sup>We shall argue in Chapter 3 that one of the main effects of distribution services on household activities is to economize on the household's use of time in purchasing tasks.

markets is quite pervasive. Many of the models in the literature that seek to explain price dispersion can be recast in these terms, despite their seemingly different approaches.<sup>12</sup>

What has not been noticed is that these equilibria usually imply differential cost shifting as well. For instance, Salop and Stiglitz (1977) seminal analysis of imperfect information requires low and high search costs consumers to generate different price equilibria; and, the implied cost shifting of information between consumers and retailers differs across these equilibria.

Generalization of the previous model to more than one distribution service is straightforward, Betancourt and Gautschi (1988). Suppose that there are two distribution services,  $D_1$  and  $D_2$ . The long run equilibrium condition becomes

$$(2.8) R = (p^* - p)/p^* = \varepsilon_1 / S_1 = \varepsilon_2 / S_2 = \varepsilon / (1 - S_Q),$$

where the subscripts 1 and 2 indicate the respective distribution service.

Equation (2.8) can be used to generate the following proposition:

*Proposition 2.* Consumer heterogeneity with respect to the valuation of different distribution services is sufficient to generate equilibria with product variety in distribution services and differential cost shifting at the same retail price, even if firms' cost functions are identical.

To illustrate the validity of this proposition, we will assume that consumers in both markets have the same constant price elasticity of demand and firms in both markets have the same cost function, given by  $C = c(v, D_1, D_2) Q^\beta$ . These assumptions ensure

---

<sup>12</sup>For an illustration in terms of rich and poor consumers with different search costs see Lal and Matutes (1989). For an illustration in which consumer heterogeneity makes it impossible to attain a first best outcome without middlemen see Biglaiser and Friedman (1997).

that the retail margin and the price in both markets will be the same, since  $R = \varepsilon/(1-\beta)$  in both cases. What differs between the two markets is at least one of the two distribution services elasticities, let us say  $\varepsilon_1(A) > \varepsilon_1(B)$  and  $\varepsilon_2 = k$  in both markets. If  $S_1$  is an increasing function of  $D_1$ , for example, both markets can be in long run equilibrium at the same retail price. If  $D_2$  is the same in both markets, this merely requires  $Q$  to be larger in market A than in market B. In any event, this generates equilibria with product variety in distribution services and differential cost shifting with respect to at least the first distribution service.

Consumer heterogeneity with respect to distribution services is an important source of equilibria with product variety in distribution services. These equilibria imply a different degree of cost shifting of distribution services. While this phenomenon has received far less attention than price dispersion in the literature, there are analyses employing different approaches that generate the same result.<sup>13</sup> For instance, Oi's (1992) location and store choice model (Section 4.4.2) requires consumer heterogeneity with respect to location and other shopping costs, which correspond to assurance of product delivery in the desired form in our terminology, to generate equilibria with product variety in distribution services and differential cost shifting for each consumer at the same full price at competing stores for the marginal consumer. The full price model is a perfectly competitive model, in contrast to the one considered here.

Extensions of the analysis to richer environments, i.e., allowing for different cost

---

<sup>13</sup>In this context we should note that the desire to allow for the role of several distribution services and the analytical difficulties in doing so lead De Palma et.al. (1994) to develop a simulation model.

functions and more than one explicit product, would expand the opportunities for retail price dispersion and product variety in distribution services to arise as characteristics of equilibria. These two consequences of the bundling of distribution services among themselves and with the explicit products sold at retail lead to the coexistence of different retail forms that we observe in the market place and will reoccur often throughout the book, e.g., in chapters 6 and 7.

## 2.5. Competition and Welfare in Retail Markets.

In this section we explore the consequences of cost shifting and bundling for competition and welfare in retail markets. We will do so in the context of a simple model which is an adaptation of several contributions in the literature. Bliss (1988) pointed out that a major problem for any retailer was to offer a consumer good enough value in the store to keep her at the store and captured this idea in terms of an indirect utility function. Betancourt and Gautschi (1993a) reformulated Bliss' idea in terms of an expenditure function and extended it by incorporating distribution services into the analysis. Betancourt and Malanoski (1999)<sup>14</sup> adapted the model for empirical analysis and noted how it generated the full price model as a special case.

Consider a representative retailer who faces the demand function in (2.2), i.e., a price setting retailer, and the cost function in (2.3). This retailer wants to choose prices and distribution services to maximize profits subject to the constraint of keeping a representative consumer patronizing the store. To keep matters simple we will focus on the single price, single distribution service situation. The problem can be specified as

---

<sup>14</sup>The model that follows borrows heavily from this last reference.

maximizing profits, defined as in (2.4), subject to the constraint that the choice of price and distribution service affects the expenditure function,  $E(p^*, p', D, Z^0)$ , of the representative consumer in the following way: It keeps her patronizing the retailer because her level of expenditures will be less than or equal to the minimum cost of attaining the same level of satisfaction at another establishment,  $E'$ .  $Z^0$  represents the optimal level of the consumption activities of the consumer.<sup>15</sup>

Formally, this is captured as follows:

$$(2.9) L = p^*Q - C(v, Q, D) - pQ + \mu[E' - E(p^*, p', D, Z^0)]$$

where  $\mu$  is a Lagrange multiplier that measures the degree of competition in a sense to be explained below. Optimal choices of retail prices and distribution services by the retailer in this setting must satisfy the following first-order conditions:

$$(2.10) p^*[1 - (1/\varepsilon)(1-\mu)] = C_Q + p$$

$$(2.11) p^*(\partial Q/\partial D) + \mu r = C_D + (C_Q + p)(\partial Q/\partial D)$$

$$(2.12) E' - E(p^*, p', D, Z^0) = 0,$$

where  $r = -(\partial E/\partial D)$  is the shadow price of distribution services or what the consumer would be willing to pay for an additional unit of distribution services if it were available in the market at an explicit price.  $\varepsilon$  is the absolute value of the price elasticity of demand for the explicit output.<sup>16</sup>

A lowering of the competitive standard faced by the optimizing retailer by one unit means a 1\$ increase in the lowest cost to a representative consumer of attaining her

---

<sup>15</sup> The expenditure function of the consumer will be discussed in greater detail in the next chapter.

<sup>16</sup> Notice that this elasticity is the reciprocal of the one in the previous section, since here we are using the ordinary demand function instead of the inverse demand function used there.

optimal level of consumption activities at an alternative establishment.  $\mu$  measures the marginal contribution to the profits of the retailer of such a lowering of the competitive standard. When  $\mu$  is zero we have the standard monopoly situation; the retailer gains no sales. When  $\mu$  is unity we have the standard competitive situation; the retailer gains all sales.<sup>17</sup> If  $\mu$  is .5, the benefit to a retailer of keeping a representative customer is half of the sales.<sup>18</sup> Thus, the value of  $\mu$  can be thought of as a direct measure of the degree of competition from other similar establishments faced by the retailer.

The power of competition from other similar establishments in retail markets can be illustrated by a special case that eliminates the economic consequences of bundling and leads to what one may label the Pareto efficient degree of cost shifting from the point of view of welfare. It generates the following proposition:

*Proposition 3:* If the marginal cost functions are independent in outputs ( $C_{DQ} = 0$ ), perfectly competitive behavior ( $\mu = 1$ ) eliminates the consequences of bundling of distribution services with the explicit items sold at retail and leads to Pareto efficient cost shifting in retail markets.

When  $\mu$  equals unity (2.10) implies that retail prices are set to equal the marginal cost of producing plus retailing the explicit items sold. This result implies for (2.11) that the level of distribution services will be set solely by the condition that the shadow price of the distribution service equals the marginal cost of producing the service. Thus,

---

<sup>17</sup>This case encompasses two different situations: Bertrand behavior where a firm behaves as if it were a perfect competitor and perfect competition proper. Note that the degree of competition in this model is exogenously determined.

<sup>18</sup>Values greater than unity violate second-order conditions. Thus, the unit interval is the relevant range.

bundling becomes irrelevant to the market outcome and the degree of cost shifting that prevails between consumers and retailers is Pareto efficient. This special case also brings out the importance of imperfectly competitive behavior in imparting on retailing its special characteristics.

When  $\mu$  departs from unity, the choice of prices by the retailer is going to be affected in general by the choice of distribution services, which is a direct consequence of the bundling of explicit items with distribution services. This observation follows directly from (2.10) and (2.11). Moreover, it also follows from these equations that the degree of cost shifting between the consumer and the retailer will in general be affected by the degree of competition from other similar establishments ( $\mu$ ) as well as the general competition for the consumer's dollar, which is measured by the price elasticity ( $\epsilon$ ).

To ascertain the consequences for welfare of retail competition, however, it is useful to look at two special cases: First, suppose that prices are given, perhaps due to government regulation. This case generates the following proposition:

*Proposition 4.* When the marginal costs with respect to explicit output are constant or increasing ( $C_{QQ} \geq 0$ ), the existence of cost shifting leads to an increase in competition from similar establishments ( $\mu$ ) increasing welfare.

When prices are given (2.10) is irrelevant. Consider the case of constant marginal costs with respect to  $Q$ , (2.11) can then be rewritten as  $M(\partial Q/\partial D) + \mu r = C_D$ , where  $M$  is a positive and constant profit margin per unit. An increase in  $\mu$  requires an increase in distribution services if  $C_D$  is increasing in  $D$ . For  $r$ , the shadow price, is decreasing in

distribution services due to the convexity of the expenditure function in the fixed input.<sup>19</sup> Moreover,  $(\partial Q/\partial D)$  must be decreasing in  $D$  or else demand would increase without bound as a result of increasing distribution services. By (2.12) an increase in  $D$  increases the welfare of the representative consumer, who can now attain the same level of utility as before at a lower cost, since the expenditure function is decreasing in  $D$ , or a higher level of utility at the same cost as before. A similar argument holds if  $C_D$  is decreasing in  $D$ , because second-order conditions in this case require marginal revenues (the LHS of 2.11) to cut marginal costs (the RHS of 2.11) from above. The importance of this result is that it contradicts the conventional wisdom stemming from Hotelling's (1929) model. The reason is that Hotelling assumed that the marginal costs of providing distribution services (location in his case) were constant at the zero level.<sup>20</sup>

Consider now a situation where the price elasticity of demand is constant but prices are not given. It allows us to establish the following proposition:

*Proposition 5:* When the marginal costs with respect to explicit output are constant or increasing ( $C_{QQ} \geq 0$ ), the existence of cost shifting leads to an increase in competition for the consumer's dollar ( $\epsilon$ ) generating ambiguous welfare results if the marginal costs of distribution services are increasing in distribution services.

Equations (2.10) and (2.11) can be rewritten as

$$(2.10)' \quad (p^* - C_Q - p) = (1/\epsilon)(1 - \mu)$$

$$(2.11)' \quad (1/\epsilon)(1 - \mu)(\partial Q/\partial D) + \mu r = C_D$$

---

<sup>19</sup>See Chapter 3 for elaboration.

<sup>20</sup>See Betancourt and Gautschi (1993a) for elaboration.

An increase in  $\varepsilon$  requires a lowering of  $p^*$  in (2.10)' under our assumptions. This increase leads to a decrease in distribution services by (2.11)' when the marginal costs of distribution services are increasing in  $D$ , since we saw above that both terms on the LHS are decreasing functions of  $D$ . From the expenditure function of the consumer in (2.12) this implies that welfare goes up due to the decrease in  $p^*$  and down due to the decrease in  $D$ . Hence, a quantitative evaluation is necessary to determine the net effect on welfare. A similar argument can be made for increases in competition through changes in  $(\mu)$ .

From the point of view of economics, the existence of cost shifting in retail markets necessitates accounting for what happens to distribution services in welfare evaluations of changes in these markets. From the point of view of marketing, the existence of cost shifting supports including distribution services in consumer's payoff functions, as proposed by Wernerfelt (1994), to evaluate marketing designs in retail markets. The model presented in this section provides one mechanism for implementing both welfare evaluations in economics and efficiency comparisons in marketing that account for the role of distribution services in retail markets.

## 2.6. Full Price Models of Retail Services.

In this section we show how the model of the previous section generates as a special case a standard model in the literature on retailing. It is a model developed by Ehrlich and Fisher (1982) to analyze the demand for advertising. They argue that if retailing is competitive consumers must face the same full price at all stores. Stores can compete by cutting prices or supplying more services but they are always subject to the constraint of the constant full price.

We can generate their model in the previous framework as follows: Interpret the distribution service ( $D$ ) as the amount of information provided by the retailer. As we saw in the analysis of Proposition 3, perfectly competitive behavior leads the first-order conditions, (2.10) and (2.11), to collapse to the following ones

$$(2.13) \quad p^* = C_Q + p$$

$$(2.14) \quad r = C_D .$$

The constraint that the full price is constant implies in our model that

$$(2.15) \quad p^* + r = C_Q + p + C_D = K$$

The full price is nothing other than the sum of the retail price and the shadow price paid by the consumer for the distribution service.  $K$  is just a constant. Our analysis brings out two restrictive features of this model. First, there is only one value of the constraint that is consistent with perfect competition in Ehrlich and Fisher's model: Namely, that value of the full price that exactly covers both marginal costs. If the firm provides so much information that the shadow price of information becomes zero to the consumer<sup>21</sup>, then the full price coincides with the retail price. As the firm lowers the amount of information provided, the value of information to the consumer increases and the retail price decreases. But, this must happen in such a way that their sum equals the retail price when the value of information is zero.

Second, and perhaps more economically relevant, the marginal cost function for information must be increasing in the amount of information or the retail market is not feasible in Ehrlich and Fisher's model. Suppose  $C_D$  is decreasing in  $D$ . When the

---

<sup>21</sup>Recall from the previous section that  $r$  is decreasing in distribution services.

constraint is imposed at the maximum level of distribution services, so that  $p^*$  is the full price because  $r$  is zero, equilibrium is possible because the constraint is not really binding. Attempts to depart from this level, however, are impossible. If the retailer decreases the level of information,  $p^*$  decreases,  $r$  increases and the marginal cost of increasing the output of explicit items increases (under the usual assumption that  $C_Q$  is increasing in  $Q$ ). The marginal cost of providing information, however, increases as you provide less information and equilibrium can not be restored. Increasing returns in the provision of information by the retailer or advertising pricing schemes incorporating quantity discounts generate decreasing marginal costs of information.

Ehrlich and Fisher's model can be useful in capturing cost shifting between consumers and retailers under decreasing returns to information. The model developed earlier, however, allows us to capture the same phenomenon without imposing this restriction.<sup>22</sup> Ehrlich and Fisher's model has been the basis for an important strand of work in the retailing literature, for example Ratchford and Stoops (1992). Furthermore, the same assumption that all consumers face a given identical full price characterizes other related contributions, for example Deacon and Sonstelie (1991). Hence, the relevance of the model in Section 2.5 is wider than the present context.

## 2.7. Measurement and Evidence.

Discussions of measurement and evidence will take place throughout the book for two reasons. First, measurement is dependent on the level of aggregation, the purpose of

---

<sup>22</sup> We will show in chapter 4 that, in general, there are compelling arguments in favor of the assumption of increasing returns or declining marginal costs in distribution services.

the analysis and the availability of data corresponding to theoretical concepts.

Measurement can take place at the level of the store, firm, sector, region or nation. In this section we will focus on measurement at the sectoral level. Second, in dealing with evidence account must be taken of the fact that distribution services, as indicated earlier in this chapter, have been acknowledged in the economics and marketing literature for many years although they have been given different names and interpretations by different authors. Evidence can also be presented at each of the levels mentioned above. At any level, however, statements in the literature referring to the service level, the marketing mix, promotion effort, etc., usually map into one or more of the five distribution services identified in Section 2.1. In this section we will focus on evidence specifically aimed at identifying these five distribution services at the sectoral level.

By sectoral level data we mean disaggregation of the retail sector of a country into its various components at some level. For instance, in the case of the U.S. the 1982 Census of Retail Trade allows disaggregation into 49 different retail sectors, 14 of which are at the four digit SIC level and the remainder at the three digit SIC level. In France and Germany one can obtain data for the same year that allows disaggregation into 50 and 52 retail subsectors, respectively. Unfortunately, the categories are not identical.<sup>23</sup> For example, the most striking difference is that in France the food sector has 14 categories while in the U.S. it has 4 and in Germany it has 5. In any event, in working with sectoral level data the choice of categories is determined, largely, by the statistical agencies

---

<sup>23</sup>For a discussion of the measurement issues that arise in international comparisons of retail subsectors see Betancourt and Gautschi (1996).

collecting the data.

The analytical objective that we will consider is understanding the determinants of retail gross margins,  $R$ , across different retail subsectors. The definition of profits in (2.4) allows this concept to be expressed as follows:

$$(2.16) R = (p^*Q - pQ) / p^*Q = C(Q, D, v) / p^*Q + \pi / p^*Q.$$

The expression on the right hand side of the first equality is unproblematic from the point of view of measurement.  $p^*Q$  is measured as the sales of the retail subsector;  $pQ$  is measured as the cost of goods sold by the retail subsector; and the ratio of their difference to sales can be used to calculate the retail gross margin. Availability of data on sales and the cost of goods sold for each subsector allows the calculation of this basic concept.

Matters get complicated, however, when we consider the expression on the right hand side of the second equality. First, the concept of profits in the definition is economic profits not accounting profits. The former are not directly observable. Second, the specification of the first term on the right hand side of the second equality depends, in general, on whether we assume quantity setting behavior, (2.1), or price setting behavior, (2.2). At this level of aggregation it is difficult to make much progress on either issue. Hence, one procedure to deal with the first difficulty is to assume that in each retail subsector monopolistic competition leads to zero economic profits.<sup>24</sup> To deal with the second difficulty we simply view the denominator in (2.16) as a revenue function.<sup>25</sup>

---

<sup>24</sup>An alternative is to specify a function of market structure variables for the second term on the RHS of (2.16). For an example of this alternative see Betancourt and Gautschi (1993b).

<sup>25</sup>In Betancourt and Gautschi (1992), for example, we pursue the implications of quantity setting and price setting in this empirical context.

Equilibrium in each subsector is characterized by the representative retailer choosing the level of distribution services and output or price that satisfies the demand of the representative consumer. The latter patronizes each subsector at different times throughout the year and demands, in general, a different combination of distribution services and output or price. In essence, this uses the time dimension to separate these retail subsector markets in the same way that the hedonic approach uses the space dimension to separate, for example, housing markets.

A semi-reduced form specification which captures the previous discussion is:

$$(2.17) R = h(Q, D; v, W) + u,$$

where  $h = C(Q, D; v) / S(Q, D; W)$ . Since  $h$  is the ratio of a cost function to a revenue function, its functional form must be nonlinear. Furthermore, the impact of a variable that appears in both on the retail margin measures the relative effect of the variable on costs and revenues.<sup>26</sup>  $u$  is a stochastic error term which will be assumed to be independent and identically distributed across the retail subsectors.

Conceptually  $Q$  is the explicit output of the retail subsector. The sales of each subsector deflated by a price index for each subsector is an appropriate measurement. Since these price indexes are not available, however, standard practice is to measure  $Q$  as sales per establishment or shop.  $D$  is a vector of distribution services. In this context a conceptually appropriate measure of accessibility of location ( $D_1$ ) is the number of establishments in each retail subsector. Fortunately, these data are usually available for

---

<sup>26</sup>Estimation of (2.17) with a cross-section of sectoral data for a particular country is facilitated because  $W$ , full income, and  $v$ , input prices, can be assumed to be the same across the retail subsectors.

each subsector.

With respect to assortment ( $D_2$ ), it is possible to construct two conceptually appropriate measures on the basis of information provided, for example, by the U.S. Census of Retail Trade. For each retail subsector one can obtain the number of establishments carrying a product line and the sales of each product line for a universe of 30 product lines.<sup>27</sup> This information allows the construction of two different indexes of assortment for each subsector. The first one is a weighted average of the number of product lines carried by a subsector, using the number of establishments in a subsector carrying a product line relative to the total number of establishments in the subsector as weights. The second one is the entropy of the distribution of sales across product lines in each subsector, i.e.,  $D_{2i} = -\sum_j (S_{ji}/S_i)\ln(S_{ji}/S_i)$ .  $j$  identifies the product line and it runs from 1 to 30 whereas  $i$  identifies the subsector. Either index measures the breadth of assortment in a retail subsector.

The remaining three distribution services identified in Section 1 are more difficult to measure at the sectoral level. Assurance of product delivery ( $D_3$ ) and information ( $D_4$ ) have several dimensions. Furthermore some of these dimensions are common to all items in an assortment and others are specific to a subset of items. In practice, assurance of product delivery is measured as the average of inventory holdings per establishment at the beginning of the year and at the end of the year and information is measured as advertising expenditures per establishment. Since the first measure fails to capture the

---

<sup>27</sup>The number of product lines in the universe varies over time for the U.S. Census and it also differs for other countries.

provision of extended hours , for example, and the second measure fails to capture the provision of information at the store, for example through in store promotions or selling effort, another empirical construct can be added to these two to capture, at least partially, these dimensions: Namely, the payroll per establishment of each subsector ( $D_6$  ). Finally, ambiance ( $D_5$ ) can be measured in the US data by the gross value of assets in building and structures per establishment for each subsector.<sup>28</sup> It is a weak measure.

In the three references cited in this section, alternative versions and extensions of (2.17) have been estimated with sectoral data for the U. S. France and Germany using a logistic functional form and alternative estimation techniques. One of two main regularities in the results is the rejection of the linear specification in favor of this nonlinear one. The other one is that the measures identified in this section, taken together, are important determinants of retail margins. Finally, we reproduce in Table 2.1 below the results for one set of estimates of (2.17) employing a logistic functional form and using nonlinear least squares as the estimation method.

Table 2.1: Estimated Coefficients.

	Q	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>5</sub>	D <sub>6</sub>
U.S.	-9.22*	-0.94*	0.12	-5.55*	82.81*	-5.66	36.41*
FRANCE	-3.12*	-0.96*	0.04	-0.21	n.a.	47.22*	19.72*
GERMANY	-1.67*	-1.22*	-0.41 <sup>∇</sup>	1.43 <sup>∇</sup>	n.a.	- 9.49	10.22*

n.a.= not available; \* = t-ratio >2.5; <sup>∇</sup> = t-ratio > 1.75.

Retail subsectors that provide greater levels of output and accessibility of location

---

<sup>28</sup>In France and Germany it can be measured as the value of new construction per establishment.

in equilibrium increase revenues by more than they increase costs and, thereby, experience lower retail margins in all three countries. Retail subsectors that provide greater levels of specific distribution services increase costs by more than they increase revenues and, thereby, experience higher retail margins in all three countries. These individual results are not only robust across the three countries, which can be seen from Table 2.1, but they are also robust to alternative versions and extensions of (2.17), which can be seen by consulting the references. They are consistent with the observed variations in these variables between, for example, supermarkets and department stores. The results on information are also robust to alternative specifications, but no data is available for France and Germany. Finally, the results on assurance and ambiance are sensitive to both country and specification. This is not surprising since these variables are poor measures of the theoretical constructs. With respect to assortment, however, the results may simply mean that its impact on costs cancels out its impact on revenues.<sup>29</sup>

By the way of a conclusion to this chapter, we consider briefly the main alternative explanation of retail margins: Namely, a mark-up model, for example Nootboom (1985).

$$(2.18) R = (OE/S)[1 + m(Q, D, X)],$$

---

<sup>29</sup>Recently, this framework has been applied to Spanish sectoral retail data for (1992) by Santos-Requejo (1996). In the case of Spain there were 25 identifiable retail sectors, mainly at the four digit level. The variables were defined and measured as indicated above whenever possible, e.g., just as in the case of France and Germany advertising expenditures data were not available for Spain. The results of the analysis are essentially the same as those for the other three countries. For example, sales per establishment and payroll per establishment are both statistically significant at the 1% level and have the same sign as reported above. While accessibility of location also has the same sign as reported above, it is not statistically significant at the 1% level. None of the other individual coefficients are statistically significant at the 1% level, which is the same result obtained above for Germany.

where OE are operating expenses excluding the shopkeeper's labor, S is sales, Q is sales per shop, D is referred to as the service bundle or product service package in the mark-up literature, and X are other variables said to affect the mark-up in various circumstances. The discussion of the service bundle in this literature refers to a subset of the same concepts that we are calling distribution services. For instance, one particular application of this model to 16 Dutch retail subsectors over the period 1976-1983, Nootboom, Kleijweg and Thurik (1988), implements the model by including the reciprocal of Q, average inventory holdings relative to sales ( $D_3/Q$  in the notation of this section),  $(OE/S)$ , and time series variables in a linear regression. The coefficient of  $(1/Q)$  is estimated to be positive with a t-ratio greater than 2.5, which is what one would expect from the results in Table 2.1 for Q in all three countries. A positive coefficient estimate with a t-ratio greater than 2.5 is also obtained for  $(D_3/Q)$ . This result is what one would expect for Holland from the result in Table 2.1 for Germany.

Summing up: the conceptualization of distribution services as outputs of retail firms and fixed inputs into the household production functions of consumers provides an attractive framework for the empirical analysis of retail gross margins; it is not rejected by the data at the sectoral level for four different countries; and it generates results that can be used to explain those of the main alternative approach to the analysis of retail gross margins employed in the literature.