

## IDENTIFYING THE COMPETITION

James LEVINSOHN

*University of Michigan, Ann Arbor, MI 48109, USA*

Robert FEENSTRA\*

*University of California, Davis, CA 95616, USA and NBER*

Received February 1989, revised version received November 1989

We propose a utility consistent method of identifying the set of competitors that a product faces. We apply the method to the 1987 U.S. new automobile market.

### 1. Introduction

A firm competes with other firms as each tries to sell its product(s) to the consumer. Much recent work in international trade theory is concerned with the degree of competition between these firms and how various policies interact with this competition. Surprisingly little attention has been focused on determining just which products compete with each other. Nonetheless, this question of ‘Who is the competition?’ is an important one.

Consider the question in the context of a market with which most readers are familiar – the U.S. automobile market. Here we would ask which automobiles compete with one another. Economic policy implications abound. Does the Japanese voluntary export restraint on automobile exports to the United States help some U.S. auto manufacturers more than others? Would an oil import fee affect one firm relatively more adversely than other firms? Finally, will an import quota on Korean automobiles benefit domestic firms or are Japanese firms the prime beneficiaries? A necessary *first* step to answering all of these questions is identifying the competition. Also, the issues presented in the above set of questions are not unique to the automobile industry. A similar set of issues and questions arise in many other industries.

\*We are grateful to Avinash Dixit for comments on an earlier draft of this paper. We are also grateful for the insightful comments of three referees. This paper was substantially revised while Levinsohn was a Visiting Fellow at the Institute for International Economic Studies, University of Stockholm. He thanks the Institute for providing an ideal work environment.

This paper develops an empirical technique, informed by economic theory, which identifies the competition when products are multi-dimensionally differentiated. Section 2 briefly discusses previous approaches to this problem. Section 3 develops the theory behind our new methodology. Section 4 discusses how one would empirically implement the theory. Section 5 provides an example of the technique. The methodology is used to identify the competitors of every 1987 model year automobile. Sensitivity analyses are also conducted. In section 6, potential problems with this new technique are discussed. We conclude with a brief discussion of possible applications of the methodology.

## 2. Previous approaches to identifying the competition

There have been two broad approaches to identifying the competition.<sup>1</sup> In the international trade literature, the typical approach to identifying the competition is to simply assume that products are differentiated by country of origin. Two very informative examples of recent empirical work on automobile trade that take this approach are Dixit (1988) and Digby, Smith and Venables (1988). Sometimes referred to as the Armington assumption, this approach is easy to implement, often meshes nicely with theoretical models in international trade, yet is usually a gross oversimplification of reality. While country of origin may be a differentiating characteristic, it is seldom the sole or most important one. Indeed, with the current proliferation of direct foreign investment, even identifying the country of origin becomes problematic (i.e. Are Nissans made in the United Kingdom European or Japanese, and are Hondas made in Ohio American or Japanese products?).

A second and often more satisfactory approach to identifying the competition is found in empirical implementations of the Hotelling model of product differentiation. Here products are arranged along a line according to quality. Any one product has at most two competitors. The location of products along the quality spectrum is accomplished with a hedonic regression. Price is regressed on salient product characteristics and the predicted price is taken as a proxy for quality. A very nice paper that adopts this approach is Bresnahan (1981).<sup>2</sup> In reality, products are multi-dimensionally differentiated and hence a given product may have more than two competitors. In the following sections we present a new technique for

<sup>1</sup>We note a third, purely statistical, technique to identify competitors. This is cluster analysis, which is discussed in Dillon and Goldstein (1984) and applied in an international trade context by Pepall (1987).

<sup>2</sup>We note that this is only the first step in a much more complex analysis of departures from marginal cost pricing. An extension of Bresnahan's complete analysis to a model with multi-dimensional product differentiation is the subject of current research by the authors.

identifying competitors that allows for multi-dimensional product differentiation. This technique, like its predecessors, makes use of a hedonic regression.

### 3. Theory

Whether firm A's product competes with firm B's product will depend upon the physical characteristics of the products and how consumers care about these characteristics. Let  $x=(x_1, x_2, \dots, x_n)$  be a vector of characteristics describing a product differentiated in  $n$  dimensions.  $R_+^n$  then is the space in which products are differentiated. The set of products available to the consumer is a discrete set  $\{x_m\}$ ,  $m=1, \dots, M$ .<sup>3</sup>  $\Theta=(\theta_1, \theta_2, \dots, \theta_n)$  represents the vector of preference parameters for an individual. Different products have different  $x$ 's and different individuals have different  $\Theta$ 's. An individual's utility is given by:

$$U(x, \Theta) + y, \tag{1}$$

where  $y$  is the quantity consumed of a numeraire good. Prices of the differentiated products are denoted by  $P_m$ . It is assumed that each consumer buys only one unit of the good. Each consumer's maximization problem can then be stated as:

$$\max_{1 \leq m \leq M} U(x_m, \Theta) - P_m. \tag{2}$$

It will be convenient to introduce the idea of a consumer's most preferred, or ideal, product [as in Lancaster (1979)]. In determining the ideal product, we hypothetically assume that all points in  $R_+^n$  are available as products. This means that we must also specify hypothetical prices for all  $x \in R_+^n$ . Prices of goods are related to the characteristics of the goods. Hence we assume that

$$P = P(x), \quad x \in R_+^n, \tag{3}$$

with  $P(x_m) = P_m$ .<sup>4</sup>

A consumer's most preferred product, denoted by  $x^*$  is given by

$$x^* = \arg \max_{x \geq 0} \{U(x, \Theta) - P(x)\}. \tag{4}$$

<sup>3</sup>The set  $\{x_m\}$  should also include alternatives to purchasing the differentiated product, such as keeping a used version. In practice, we shall not include these alternatives as data.

<sup>4</sup>In other words, we are assuming that the price function (3) fits the price and characteristics data on the  $M$  models exactly. In practice, we expect that  $P_m$  will differ from  $P(x_m)$  by some error, as discussed in section 5.

The first-order condition implied by (4) is given by:

$$\frac{\partial U(x^*, \Theta)}{\partial x} = \frac{\partial P(x^*)}{\partial x}. \quad (5)$$

For many functional forms of (1), condition (5) may be inverted to solve for the unobservable taste parameters.<sup>5</sup> This implies a mapping between tastes and most preferred products given by:

$$\Theta = \Psi(x^*). \quad (6)$$

Suppose  $F(\Theta)$  denotes the cumulative distribution function of consumers with taste parameters less than or equal to  $\Theta$ . Then we can use (6) to define the distribution of consumers over ideal products as  $G(x^*) \equiv F[\Psi(x^*)]$ . We assume that the support of the distribution  $G$  is a convex set which includes all available models.<sup>6</sup>

Now define,

$$S(x, x^*) \equiv U(x, \Psi(x^*)) - P(x). \quad (7)$$

Eq. (7) defines a surplus function  $S(x, x^*)$ . This function gives the surplus associated with product  $x$  when the ideal product is  $x^*$ . Surplus is maximized at  $x = x^*$ . Likewise, as  $x$  differs from  $x^*$  in any of the  $n$  dimensions in which the products are differentiated, surplus declines.<sup>7</sup> Eq. (7) also defines iso-surplus contours conditional on a most preferred product,  $x^*$ . Surplus is constant along a contour and declines as contours become further from  $x^*$ . The shape of these contours will depend upon the functional forms chosen for (1) and (3).

Now the consumer's problem, (2), may be restated as:

<sup>5</sup>From Gale and Nikaido (1965), a sufficient condition to globally invert (5) obtaining (6) is that the principal minors of  $U_{x,\theta}$  be positive for all  $x > 0$  and  $\theta > 0$ . This will be satisfied by the functional form used in section 3.

<sup>6</sup>This assumption is made to ensure that there exists at least one consumer with every relevant ideal variety. See also section 6.

<sup>7</sup>To see this, consider the second-order approximation  $S(x, x^*) \approx S(x^*, x^*) + S_x(x^*, x^*)(x - x^*) + (1/2)(x - x^*)'S_{xx}(x^*, x^*)(x - x^*)$ , where, in this footnote only, subscripts denote partial derivatives. From (5) and (7), we see that  $S_x(x^*, x^*) = 0$  and  $S_{xx}(x^*, x^*) = [U_{xx}(x^*, \Theta) - P_{xx}(x^*)]$ , which is negative definite from the second-order conditions for (4). (Note, though, that this is a local property only.) It follows that  $S(x, x^*) \approx S(x^*, x^*) + (1/2)(x - x^*)'[U_{xx}(x^*, \Theta) - P_{xx}(x^*)](x - x^*)$ , where the latter term is negative for  $x \neq x^*$ .

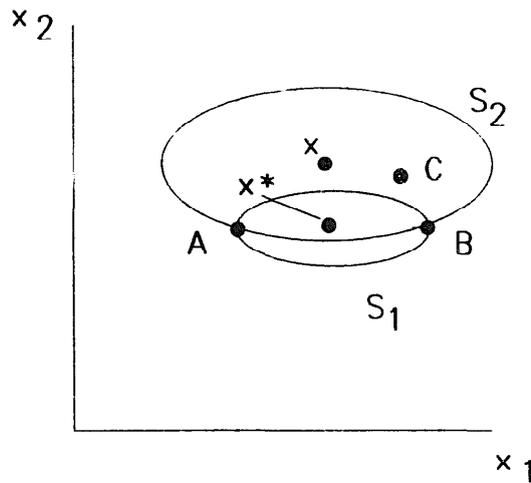


Fig. 1

$$\max_{1 \leq m \leq M} S(x_m, x^*) = U(x, \Psi(x^*)) - P_m. \tag{8}$$

It is clear that (8) is simply a rewritten version of problem (2), in the sense that model  $m^*$  will solve both problems if  $\Theta = \Psi(x^*)$ .

We propose the following definition of competitors.

**Definition.** Products A and B are competitors if there exists  $x^* \in R_+^n$  such that  $S(x_a, x^*) = S(x_b, x^*) \geq S(x_c, x^*)$ ,  $\forall$  models C.

This definition states that if A and B are competitors, then there exists a consumer indifferent between A and B who prefers those two models to all other models. There are, though, an infinite number of ideal varieties,  $x^*$ , that still leave  $S(x_a, x^*) = S(x_b, x^*)$ . We need only find one point  $x^*$  satisfying the above condition to call A and B competitors.

Fig. 1 illustrates our definition. Points A, B, and C represent available products differentiated in characteristics  $x_1$  and  $x_2$ .  $S_1$  represents the surplus contour of an individual with an ideal product  $x^*$  who is indifferent between A and B (since A and B lie on the same iso-surplus contour),  $S_2$  represents the surplus contour of an individual with an ideal product  $x$  who is also indifferent between A and B. In fig. 1, models A and B are competitors since  $S(x_a, x^*) [= S(x_b, x^*)] > S(x_c, x^*)$  as drawn. The fact that  $S(x_a, x) [= S(x_b, x)] < S(x_c, x)$  means that a consumer with ideal product  $x$  would buy model C rather than A or B.

The above definition has several appealing qualities. These include: (a) as in a Hotelling model, relations between competitors may be intransitive; (b)

the definition is symmetric (if A is a competitor of B, B is necessarily a competitor of A); and (c) the definition is equivalent to defining two products as competitors if and only if their cross-elasticity of demand is positive. This equivalence can be seen as follows.<sup>8</sup>

Suppose that A and B are competitors by the above definition. Then there is a consumer with ideal variety  $x^*$  such that  $S(x_a, x^*) = S(x_b, x^*) \geq S(x_c, x^*)$  for all models C. Ignoring borderline cases, treat the latter inequality as strict. Now raise the price of model A by a small amount. This increases  $P_a$  and lowers  $S(x_a, x^*)$  in (8).<sup>9</sup> For a sufficiently small increase in  $P_a$ , we will then have  $S(x_b, x^*) > S(x_a, x^*) > S(x_c, x^*)$ , and so the consumer with ideal  $x^*$  will now purchase model B. Thus, when products are competitors by our definition, a slight increase in the price of one will shift consumers towards purchasing the other. It follows that the cross-price elasticity of demand is positive. Conversely, if two products are not competitors by our definition, then there are no consumers indifferent between them, and the cross-price elasticity of demand is zero.

#### 4. Implementing the theory

The first step towards implementing the theory requires imposing functional forms on the utility function (1) and on the price function (3). While there is not a clearly right or wrong functional form for utility, some functional forms are better than others. One requirement is that the number of taste parameters,  $\theta_i$ , equal the number of characteristics,  $x_i$ . In addition, it is desirable to be able to vary the concavity of the utility function in a parametric manner. These goals are achieved by using a constant elasticity of substitution functional form:

$$U(x, \theta) = \sum_{i=1}^n \frac{\theta_i}{\delta} (x_i^\delta - 1). \quad (9)$$

The elasticity of substitution between characteristics is  $\sigma = 1/(1 - \delta)$  with  $\delta \leq 1$ .<sup>10</sup>

It will be important for empirical work that the price function (3) fit the data well. With this requirement in mind, empirical work often imposes (or tests) a log-linear functional form on the price function. Hence:

$$P(x) = \exp(\alpha + \beta'x), \quad (10)$$

<sup>8</sup>A rigorous definition of demand, and its relation to the set of competitors for each product, is provided in Feenstra and Levinsohn (1989).

<sup>9</sup>When raising  $P_a$  in (8), we hold the function  $\Psi(x^*)$  constant.

<sup>10</sup>As  $\delta \rightarrow 0$ , the utility function becomes  $U(x, \theta) = \sum_{i=1}^n \theta_i \ln x_i$ .

where  $\alpha$  and  $\beta = (\beta_1, \beta_2, \dots, \beta_n) > 0$  are parameters.

Given these functional forms, the mapping (6) between tastes and most preferred products becomes:

$$\theta_i = (x_i^*)^{1-\delta} \beta_i \exp(\beta' x^* + \alpha). \tag{11}$$

The surplus function is:

$$S(x, x^*) = \exp(\beta' x^* + \alpha) \sum_{i=1}^n \frac{\beta_i}{\delta} (x_i^*)^{1-\delta} (x_i^\delta - 1) - \exp(\beta' x + \alpha). \tag{12}$$

The iso-surplus contours implicitly defined by (12) resemble ellipsoids but are not symmetric. Rather, they are slightly egg-shaped.

Before the surplus function can be used to identify competitors, it must first be parameterized. We will obtain estimates of the  $\beta_i$ 's by estimating (the log of) (10). The parameter  $\delta$  relating to the elasticity of substitution between characteristics is simply posited. The robustness of results to the choice of  $\delta$  is investigated in subsequent sensitivity analyses. The characteristics of an available model, the  $x_i$ 's, are data.

Note that the definition of competitors, using (12), is invariant to the units in which characteristics are measured. Suppose, for example, that units of measurement were redefined such that  $x_i$  is now scaled by a factor  $\lambda$ . This implies that the  $\beta$  estimated from (10) is multiplied by  $1/\lambda$ . Then simple calculations show that the surplus function becomes:

$$S(\lambda x, \lambda x^*) = \exp(\beta' x^* + \alpha) \sum_{i=1}^n \frac{\beta_i}{\delta} (x_i^*)^{1-\delta} (x_i^\delta - \lambda^{-\delta}) - \exp(\beta' x + \alpha). \tag{12'}$$

It is straightforward to show that (12) and (12') lead to identical sets of competitors. The reason is that when making the comparison,  $S(x_a, x^*) = S(x_b, x^*) \geq S(x_c, x^*)$ , as in our definition. The term involving  $\lambda^{-\delta}$  in (12') will cancel out. The remaining terms in (12') are identical to those in (12), so identical sets of competitors are obtained.

Given two potential competitors, products A and B, the next step is to identify the consumers with preferences  $\Theta$ , and hence a most preferred product  $x^*$  such that  $S(x_a, x^*) = S(x_b, x^*)$ . With  $x^*$  identified, the surplus function is fully parameterized. As discussed in the previous section, there are an infinite number of points  $x^*$  such that  $S(x_a, x^*) = S(x_b, x^*)$ , and in principle we need to check each of them to see whether our definition holds. In

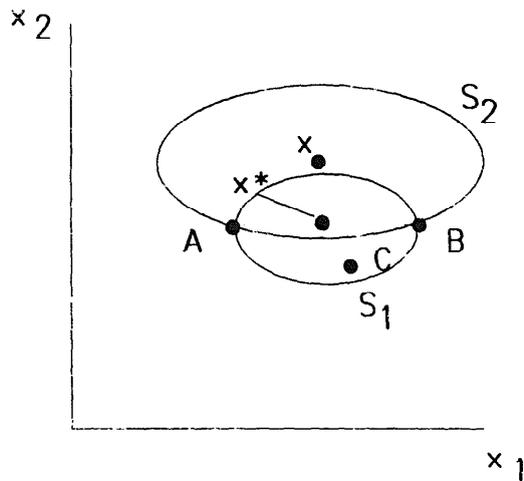


Fig. 2

practice, we shall simplify the task by only considering points  $x^*$  such that  $x^* = \Omega x_a + (1 - \Omega)x_b$ , with  $0 \leq \Omega \leq 1$ . That is, we restrict attention to ideal products  $x^*$  which lie on a line segment between  $x_a$  and  $x_b$ . In many cases, such as illustrated in fig. 1, this approach will correctly identify whether A and B are competitors. However, since we have restricted the search for  $x^*$ , it is certainly possible that we will *omit* some models as competitors, as illustrated in fig. 2.

Fig. 2 is the same as fig. 1, except that the position of model C has changed. Ideal product  $x^*$  satisfies  $S(x_a, x^*) = S(x_b, x^*)$  and  $x^* = \Omega x_a + (1 - \Omega)x_b$ . Since  $S(x_a, x^*) [= S(x_b, x^*)] < S(x_c, x^*)$ , we would reject models A and B as competitors by only considering  $x^*$  lying on a line segment lying between  $x_a$  and  $x_b$ . However, we also see that  $S(x_a, x) [= S(x_b, x)] > S(x_c, x)$ , so models A and B are competitors by our definition. Summing up, our method can reject models as competitors when they really are, but it clearly can never accept models as competitors when they are not. In section 6 we discuss how serious this limitation may be.

With many models A, B, C, D, ... available, we calculate  $x^*$  on a line segment between each two models with equal surplus obtained from the two. We then check whether greater surplus is obtained from any other model. If so, then the models are (possibly incorrectly) rejected as competitors; if not the two models are accepted as competitors. The results from this calculation are reported next.

## 5. Competitors in automobiles

In this section we identify the competitors to the 136 models which make

up the 1987 new car market in the United States.<sup>11</sup> The automobile market provides an especially nice test of our methodology. It is a market with which most readers are familiar, hence the plausibility of our results is readily examined. It is also a market with many differentiated products and plentiful data.

We first choose the dimensions in which products are differentiated. Data were available on 12 characteristics of every model.<sup>12</sup> We choose the most relevant characteristics by applying nested hypothesis testing to the OLS estimation of the price function (10). We address possible problems with the use of nested hypothesis testing in the sensitivity analyses.

The estimated price function is

$$\begin{aligned} \text{LOGPRICE} = & -0.685 + 0.000119WT + 0.121CARBI + 0.0030TORQ \\ & (0.519) (0.000059) \quad (0.051) \quad (0.0007) \\ & + 0.156PSI + 0.407AIRI + 0.252FORI + 53.8INVHT, \\ & (0.053) \quad (0.053) \quad (0.045) \quad (25.6) \\ R^2 = & 0.794, \quad 136 \text{ observations.} \end{aligned} \quad (13)$$

Of the 12 differentiating characteristics on which data were available, nested hypothesis testing resulted in a characteristics space differentiated in seven dimensions. These are weight of the vehicle (*WT*), engine torque (*TORQ*), the inverse of the height of the vehicle (*INVHT*), whether the auto had fuel injection (*CARBI*), air conditioning (*AIRI*), and power steering (*PSI*) as standard equipment, and whether the auto was foreign (*FORI*).<sup>13</sup> All variables refer to the base model of a product.

The use of binary variables in the context of identifying competition deserves special note. We view the variables *PSI* and *AIRI* as proxies for increasing degrees of luxury. Similarly, the *FORI* may proxy for perceived quality or longevity. In these contexts, a most preferred product may well

<sup>11</sup>When a model is produced by two divisions of the same corporation and the models only differ cosmetically, such as the Ford Escort and the Mercury Lynx, only one of the models is used in the sample.

<sup>12</sup>Please refer to the data appendix for a complete description of the data set.

<sup>13</sup>The dummy variables *CARBI*, *PSI*, and *AIRI* took the value of 2 if the feature was standard equipment and 1 if not. *FORI* took the value of 2 if the auto was produced abroad and 1 if domestically produced. (Models which were produced both in the United States and abroad are considered foreign models.) This differs from the usual 0-1 convention, but is necessary since some dummy variables are raised to negative powers in calculating the surplus function. This departure from convention only changes the constant in the regression and has no effect on the results.

have a value for these variables that is between the all-or-nothing choice imposed by available products.

Characteristics such as horsepower and miles per gallon of gasoline, items many consumers deem important, were not statistically significant<sup>14</sup> in the price function. This is because these characteristics are spanned by linear combinations of included variables such as weight, inverse of height, and torque. Including the statistically insignificant characteristic of the price function (13) would have posed two problems. First, inclusion would have induced multicollinearity. This is a relatively minor problem as estimates are still unbiased, and we only use the parameter estimates, not their estimated standard errors, in the surplus function.<sup>15</sup> More importantly, each characteristic included in the price function represents a dimension in which one must search for potential competitors. With too many characteristics, the search becomes computationally burdensome.

Estimated coefficients in the price function are used to parameterize the surplus function (12). The only unidentified parameter in the surplus function is  $\delta$  which is related to the elasticity of substitution between characteristics. We are unable to econometrically identify  $\delta$  with our data. Rather, we set  $\delta$  equal to a variety of plausible values and test the robustness of our results.<sup>16</sup> Setting  $\delta = -3$  for the base case scenario, competitors to every 1987 automobile model were identified using the procedure described above. The results are given in table 1. Table 1 tells one, for example, that model 2, the Acura Legend, has as competitors models: 4 – the Alfa Romeo Milano, 8 – the Sterling, 40 – the Nissan Maxima, and 69 – the Volvo 740. The number of competitors to a model varies from one (the Chevrolet Corvette) to sixteen (the Renault Medallion). With few exceptions, the results accord well with intuition.

If only because the theoretical literature so often assumes product differentiation only by country of origin, it is worth noting that the results in table 1 are broadly inconsistent with this assumption. For example, two Japanese products, the Acura Legend and the Toyota Tercel, do not have a single competitor in common. Rather, the Legend's competitors are fairly large powerful European models, while the small Tercel's competitors are all Japanese models.

The results in table 1 are the product of several rather ad hoc assumptions made in order to empirically implement our theory. It would be helpful to test whether the results obtained are robust to these assumptions. In this

<sup>14</sup>'Statistically significant' refers to an estimate being statistically significantly different from zero at the 95 percent level.

<sup>15</sup>Were multicollinearity too severe, though, the design matrix might be so ill-conditioned that inversion difficulties arise. In addition, negative estimated values for  $\beta_i$  can lead the surplus function (12) to violate the properties in footnote 5.

<sup>16</sup>This procedure was used in a different context by Feenstra (1988). Note that lower (negative) values for  $\delta$  lead to utility surplus functions which are more concave.





Table 1 (continued)

Model no.	Model name	Model numbers of competitors									
100	Shadow	47	81	95	118	125					
101	Aries	26	98	107	128	129					
102	Daytona	57	59	75	81	89	108	128			
103	Lancer	39	47	81	104	108	109	119			
104	600	47	81	103	126						
105	Diplomat	72	82	83	86	93	97	127			
106	Escort	65	88	98	115	117	135				
107	Tempo	66	81	91	101	125	129	136			
108	Mustang	47	57	75	81	96	102	103			
109	Taurus	39	43	47	57	96	103	119			
110	TBird	29	30	46	78	79	80	86	121		
111	LTD	12	83	86	94	113	127				
112	Continental	12	113								
113	TownCar	12	83	111	112						
114	Sentra	21	25	56	62	65	98	99			
115	Nova	20	22	87	98	106	135				
116	FX16	17	31	38	52	53	54	64	1	136	
117	Firenza	67	73	88	91	98	106				
118	Calais	47	74	75	100	125					
119	Ciera	43	47	76	90	96	103	109			
120	Supreme	132									
121	Delta88	29	78	93	110	134					
122	Olds98	29	84	85	134						
123	Toronado	61	80	84	86						
124	Turismo	21	77	88	99	114	129				
125	Sundance	47	60	81	100	107	118				
126	Caravelle	95	104								
127	GranFury	72	82	83	86	93	105				
128	Fiero	3	27	101	102	129					
129	Sunbird	26	52	88	91	98	101	107	124	128	136
130	GrandAm	75	79	89	133						
131	Firebird	15	69	78	81	89	92				
132	GrandPrix	15	72	77	120						
133	6000	43	47	76	89	90	130				
134	Bonneville	29	82	93	121	122					
135	Alliance	9	17	31	81	87	88	106	115		
136	GTA	66	88	91	107	116	129				

context, Leamer (1985) has argued persuasively for sensitivity analysis. Two types of sensitivity analysis were performed to check the robustness of the results in table 1. First,  $\delta$  was set equal to  $-1.0$  and then to  $-0.01$ . This experiment varies the degree of concavity of the utility function. Each time, the analysis was repeated.

Second, the order in which hypotheses are tested when conducting nested hypothesis testing may affect the outcome of the tests. For example, by slightly changing the criteria used for the selection of variables in eq. (13) and changing the order of the nested hypothesis tests, one can arrive at a hedonic regression that includes horsepower (*HP*) instead of torque. Recog-

Table 2  
Sensitivity analyses:  $S$  = a summary statistic  
for competing models.

	Mean	Standard dev.
Base case ( $\delta = -3.0$ )	1.0809	0.2509
$\delta = -1.0$	1.0597	0.2842
$\delta = -0.01$	1.0668	0.2610
<i>HP</i> instead of <i>TORQ</i>	1.0810	0.3015

nizing this, we used engine horsepower in the price function instead of using torque, and repeated the analysis.

Robustness of the results depends on the question being posed. If one is solely interested in whether model A competes with model B, one finds that about three out of every four competitors remain when either  $\delta$  changes or *HP* is used. On the other hand, one may wish to characterize aspects of the entire group of competitors. One may wish to ask, for example, what happens to the average price of a model's competitors.

There are several plausible statistics that would adequately represent how stable the set of competitors that a model faces is. Denote the set of competitors that model  $m$  faces by  $I_m$  with  $N_m$  elements. One such summary statistic,  $S$ , is:

$$S = \frac{1}{M} \sum_{m=1}^M \left[ \frac{(1/N_m) \sum_{n \in I_m} P_n}{P_m} \right]. \quad (14)$$

The mean price of a model's competitors is in the numerator of the term in brackets. This mean is normalized by the model's own price. Finally, the term in brackets is averaged over all models. For example, if  $S = 1.0$ , on average a model has the same price as its competitors. In table 2 we show how this summary statistic varies as we vary  $\delta$  and replace torque with horsepower. The robustness of the results is illustrated by how little  $S$  varies across sensitivity analyses. Given the standard deviation of  $S$  reported in the second column of the table, it is clear that the results are quite robust (if one cares about the average price of a model's competitors).<sup>17</sup>

In general, then, if one wishes only to determine if model X is a competitor of model Y, answers may change with specification changes. If, on the other hand, one wants to know about the (characteristics of the) group of

<sup>17</sup>More detailed results of the sensitivity analyses are available on request.

competitors a product faces, results appear quite robust. Robustness of the results will in the final analysis depend on just what issue is being analyzed.

## 6. Caveats

We believe the methodology discussed in this paper is a valuable tool and can be profitably applied to many economic problems. However, our analysis has a number of limitations. In this section we discuss four such limitations.

First, when calculating surplus in (12) we have subtracted the predicted price from the hedonic regression (13) rather than the actual price. This means that we are treating the error term from the hedonic regression as reflecting only unmeasured characteristics of a model, which yield utility exactly equal to the dollar error. It follows that surplus is properly calculated as the utility from measured characteristics minus the predicted price, as in (12).

An alternative approach we considered was to treat the error in the hedonic regression as reflecting pure price markups, with no unmeasured characteristics. In this case, actual rather than predicted prices are used in (12). However, this approach fails in practice, because we find that certain models with actual prices much greater than predicted prices are never purchased. That is to say, a grid search over  $\Theta$  shows that every consumer would avoid the high prices of certain models, and choose another model with similar characteristics. In future research it would be desirable to assume that the errors in (13) reflect some combination of unmeasured characteristics and pure price markups.

Secondly, because most preferred products lie on lines drawn between available products, the methodology may omit some models as competitors. In particular, it is unable to account for the preferences of consumers whose ideal product lies outside the convex hull of all available products.<sup>18</sup> This is unlikely to be a problem in a market with minimal entry barriers facing new products. Were there a high density of consumers in a part of characteristics space outside the convex hull of available models, such a market niche would likely be a profitable one. Given the actual absence of available models, we conclude that there are not likely to be many consumers whose preferences lie outside the convex hull of available products. While this caveat, then, may not apply to the automobile market with its myriad products, it may well apply to the super-computer market.

Thirdly, we define the competition without using any actual sales data. Rather, we simply estimate a hedonic regression. We are able to define the competition in the absence of specific sales data by using a structural model

<sup>18</sup>In fig. 2, the convex hull of available products is the triangle with vertices at A, B, and C. Our method fails to identify A and B as competitors because the ideal product  $x$  lies outside the convex hull.

in which estimated implicit prices of characteristics are interpreted as marginal utilities. By assuming a specific functional form, we are then able to obtain the utility function. The choice of the utility function is not innocuous. The CES functional form of (9) is often used in the characteristics approach to product differentiation, allows us to vary  $\delta$  for purposes of sensitivity analysis, and still yields an invertible first-order condition to (4).<sup>19</sup> In current research we have obtained similar sets of competitors by using a quadratic surplus function in place of (12), but we have not yet experimented with other functional forms.

Fourthly, our procedure is not conducive to formal hypothesis testing of which products are and which products are not competitors. Rather, the 'confidence' one might place in the list of competitors is investigated by conducting sensitivity analyses. Because of this aspect of our method, sensitivity analyses seem essential.

## **7. Potential applications**

Product differentiation has recently been a much researched topic in the international trade and the industrial organization literatures. Almost all of this research has been theoretical. A wide range of these issues may be empirically investigated using, usually as a first step, the methodology described above. Levinsohn (1988) estimated an ad hoc demand system in which identification of competitors was used to impose cross-price elasticity zero restrictions. Work on estimating a utility consistent demand system for differentiated products as well as work estimating oligopoly pricing of differentiated products is in progress by the authors.

Anti-trust analysis is often directed at ascertaining whether firm mergers will give rise to market power in a particular segment of the market. The answer often depends on how the market is defined. For example, whether one includes international competitors may affect the answer. The method of identifying the competition is a natural tool for the job. Similarly, one could easily investigate the dynamic competitive effects of government policies such as domestic taxes and subsidies, bail-outs, tariffs, and quotas by analyzing how competitors change over time in response to the policies.

The methodology also has natural marketing applications. Given the characteristics of a potential entrant, it is straightforward to determine with which products the potential entrant would compete.

These are but a few examples. The methodology presented in this paper is a first attempt at devising a much needed empirical tool. We hope the methodology will facilitate work in international trade, industrial organiza-

<sup>19</sup>See also footnote 5.

tion, and public economics. We also hope others are stimulated to improve on the methodology itself.

### Data appendix

The data used are from the 1987 *Automotive News Market Data Book*. The entire data set is available on floppy disk from the authors on request. Collected variables (and their units of measurement) are Overall Length (inches), Overall Width (inches), Overall Height (inches), Curb weight (pounds), Engine Displacement (cubic inches), Carburation (2 in fuel injected, 1 otherwise), Net Horsepower, Net Torque (foot pounds), Power Steering (1 if not standard, 2 if standard), Power Brakes (1 if not standard, 2 if standard), Air Conditioning (1 if not standard, 2 if standard), Foreign (1 if domestic, 2 if foreign), and List Price (dollars). For some but not all models, the EPA miles per gallon figures are reported.

The source code of the FORTRAN 77 programs which implemented the identification of competitors and the sensitivity analyses are also available from the authors on request.

### References

- Bresnahan, T, 1981, Departures from marginal-cost pricing in the American automobile industry, *Journal of Econometrics* 17, 201–227.
- Crain Communications, 1987, 1987 Automotive news market data book (Crain, Detroit, MI).
- Digby, C., A. Smith and A. Venables, 1988, Counting the cost of voluntary export restrictions in the European car market, Centre for Economic Policy Research Discussion Paper no. 249.
- Dillon, W. and M. Goldstein, 1984, *Multivariate analysis: Methods and applications*, Wiley series in probability and mathematical statistics (Wiley, New York).
- Dixit, A. 1988, Optimal trade and industrial policies for the U.S. automobile industry, in: R. Feenstra, ed., *Empirical methods for international trade* (MIT Press, Cambridge, MA).
- Feenstra, R., 1988, Gains from trade in differentiated products: Japanese compact trucks, in: R. Feenstra, ed., *Empirical methods for international trade* (MIT Press, Cambridge, MA).
- Feenstra, R. and J. Levinsohn, 1989, Distance, demand, and oligopoly pricing, Mimeo.
- Gale, D. and H. Nikaido, 1965, The Jacobian matrix and global univalence in mappings, *Mathematische Annalen* 159, 81–93.
- Hotelling, H., 1921, Stability in competition, *Economic Journal* 39, 41–57.
- Lancaster, K., 1979, *Variety, equity, and efficiency* (Columbia University Press, New York).
- Leamer, E., 1985, Sensitivity analyses would help, *American Economic Review* 75, 308–313.
- Levinsohn, J., 1988, Empirics of taxes on differentiated products: The case of tariffs in the U.S. automobile industry, in: Robert Baldwin, ed., *Trade policy issues and empirical analysis* (University of Chicago Press, Chicago, IL).
- Pepall, L., 1987, A clustering analysis of the effects of import penetration on product variety, Mimeo. (Tufts University, Medford, MA).