Appendix

1 Deriving the aggregate demands from the individual indirect–utility functions

Suppose that there are n brands of a differentiated product, each of which is produced by one of K firms.[1] Let $\tilde{p}_i$ be the nominal price of the ith brand, $i = 1, \ldots, n$. All other goods are aggregated into an outside good that is sold at a nominal price p_0. The market for the differentiated product is assumed to be imperfectly competitive, whereas the outside good is competitively supplied.

The hth individual consumes a vector $q_h = (q_{h1}, \ldots, q_{hn})^T$ of the differentiated product, with $q_{hi} \geq 0, i = 1, \ldots, n$, and q_{h0} of the outside good, with $q_{h0} > 0, h = 1, \ldots, H$.[2] Furthermore, this individual has nominal income $\tilde{y}_h$ and indirect–utility function $\tilde{u}_h(p_0, \tilde{p}, \tilde{y}_h)$, where $\tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_n)^T$.

Not knowing the functional form of $\tilde{u}_h$, we approximate it with a flexible functional form. We choose a normalized quadratic (Berndt, Fuss, and Waverman 1977 and McFadden 1978b), where prices and incomes are normalized or divided by p_0. Specifically, $\tilde{u}_h(p_0, \tilde{p}, \tilde{y}_h) \approx p_0 u_h(1, p, y_h)$, where $p = p_0^{-1}\tilde{p}, y_h = p_0^{-1}\tilde{y}_h$,

$$u_h(1, p, y_h) = -\left[a_{h0} + a_h^T p - p_0 y_h(\gamma_0 + \gamma^T p) + \frac{p_0}{2}\, p^T B_h p\right], \tag{1}$$

and each B_h is an arbitrary $n \times n$ symmetric, negative–semidefinite matrix.

This utility function is flexible in prices. In other words, it is a second-order approximation that places no restrictions on substitution possibilities between brands of the differentiated product. Moreover, the function is in Gorman polar form and can therefore be aggregated to obtain brand–level demands.[3] Furthermore, unlike many other models (e.g., random–utility models), one need not specify the distribution of unobserved consumer heterogeneity in order to aggregate individual demands.

Unfortunately, the matrix $B_h = [b_{hij}]$ alone has $n(n+1)/2$ parameters, and equation (1) must be simplified. We assume that a_{hi} and b_{hii}, $i = 1, \ldots n$, the linear and quadratic coefficients of p_i in the utility function, are linear functions of the characteristics x_i of brand i. For example, when the product is beer, the characteristics

[1] Firms can produce more than one brand, but no brand is produced by more than one firm.

[2] If $q_{hi} = 0$, the income effect for that good must be zero (i.e., $\gamma_i = 0$ below).

[3] See Gorman (1953, 1961) or Blackorby, Primont, and Russell (1978) for discussions of the conditions that are required for consistent aggregation across households.

might be the brand's alcohol content, product type (e.g., lager, ale, or stout), and brewer identity. Moreover, those functions can differ by individual, $a_{hi} = a_h(x_i)$ and $b_{hii} = b_h(x_i)$. In contrast, the off–diagonal elements of B_h, which are the coefficients of the interaction terms, $p_i \times p_j$ for $j \neq i$, are assumed to be functions of a vector of measures, d_{ij}, of the distance between brands in some set of metrics. For example, when the product is beer, the measures of distance, or its inverse closeness, might be alcoholic-content proximity and dummy variables that indicate whether the brands belong to the same product type (e.g., whether both are stouts) and whether they are brewed by the same firm. In addition, one can construct distance indicators that have been used by others, such as the common–market–boundary measure of Feenstra and Lev insohn (1995). The functions of distance can also differ by consumer, $b_{hij} = g_h(d_{ij})$.

By Roy's identity, individual demands for the differentiated product are

$$q_{hi} = \frac{a_{hi} + \Sigma_j b_{hij} p_j - \gamma_i y_h}{p_0(\gamma_0 + \gamma^T p)}. \tag{2}$$

In equation (2), $p_0(\gamma_0 + \gamma^T p) = \gamma_0 p_0 + \gamma^T \tilde{p}$ is a price index that can, without loss of generality, be set equal to one in a cross section or very short time series. After this normalization, aggregate product demands become

$$q_i = \sum_h a_{hi} + \sum_j \left(\sum_h b_{hij} \right) p_j - \gamma_i \left(\sum_h y_h \right) = a_i + \sum_j b_{ij} p_j - \gamma_i y, \tag{3}$$

where $a_i = \sum_h a_{hi}$, $b_{ij} = \sum_h b_{hij}$, and $y = \sum_h y_h$ is aggregate income.

Equation (3) shows that the brand–level demand intercept, which determines the size of the brand's market, depends on product and market characteristics, x_i and y.[4] This assumption transforms the model from one in which consumers demand brands into one in which they demand the characteristics that are embodied in those brands, as in a hedonic study. If the number of characteristics is less than the number of brands, the dimensionality of the problem is reduced.

The diagonal elements , b_{ii}, which determine the own–price elasticities, also depend on the characteristics, the hypothesis being that, for example, the demand for high–alcohol beers might be systematically less elastic than that for low. Off–diagonal elements, $b_{ij}, j \neq i$, in contrast, which determine substitutability between brands,

[4] A market is a regional/time-period pair with zero cross-price elasticities across markets. More generally, in addition to per–capita income, market characteristics might include, for example, population and the number of firms that supply the product in the region. The form of the utility function, however, (i.e., the fact that it is in Gorman polar form) implies that only average consumer characteristics matter.

depend on distance measures, the hypothesis being that, for example, brands that have similar alcohol contents might be closer substitutes.

Let X be the matrix of observed brand and market variables with typical row $X_i = (x_i^T, y)^T$. If there are also unobserved brand and regional characteristics u, (3) can be written in matrix notation as

$$q = \alpha + X\beta + Bp + u, \tag{4}$$

where α is a vector of intercepts that we treat as random effects, and β is a vector of parameters that must be estimated.

The matrix $B = (b_{ij})$ has two parts: b_{ii} is a parametric function of X_i, and $b_{ij} = g(d_{ij})$, $i \neq j$, is a function of the measures of distance between brands i and j. As we are interested in placing as little structure as possible on substitution patterns, we estimate $g(.)$ by semiparametric methods.

Finally, the random variable u, which captures the influence of unobserved product and market variables, can be heteroskedastic and correlated across observations. We assume, however, that the unobserved characteristics, u, are mean independent of the observed characteristics, X, $E[u_i|X] = 0$. Whereas this assumption is problematic, it is standard in the literature.[5] Moreover, it can be tested, as we do below.

2 Semiparametric Estimation

2.1 Outline

The semiparametric estimation method used here is that of Pinkse, Slade and Brett (1998). The method allows the arguments of the function g to be a combination of discrete and continuous measures. The discrete measures can only take finitely many different values and the function g should be continuous in each of the continuous

measures, which preferably have a compact support (as we assume is the case here).

Suppose that there are C permutations of the values of the discrete distance measures. For each category we could allow for a different continuous function of the continuous distance measures. In the semiparametric specification discussed in section 7.2, however, g_{ij} is assumed zero whenever observations i and j correspond to different regions, time periods or product types, which leaves us with just one continuous function to be estimated, say g^*.

[5] For a discussion of the difficulties involved in relaxing this assumption, see Berry (1994).

The function g^* can be written as an infinite order linear combination of functionals chosen by us and unknown coefficients. We use a Fourier series expansion, meaning that we express g^* (on $[0, 2\pi]$) in terms of the series $1, \cos x, \sin x, \cos(2x), \sin(2x), \ldots$. We cannot estimate all infinitely many unknown coefficients, but limit ourselves (in the application) to the first three or five. The number of expansion terms can be chosen by graphical inspection (which we do) or by an automatic selection method. We did not find much variation in the results across reasonable choices of the number of expansion coefficients estimated. An estimate of g^* can then be constructed as a linear combination of coefficient estimates and the corresponding functionals.

More concretely, the model in (??) can be rewritten as

$$q_{irt} = b_{irt}p_{irt} + n_{\mathcal{G}_{irt}}^{-1} \sum_{j \in \mathcal{G}_{irt}} g^*(d_{ijrt}^C)p_{jrt} + \beta^T x_{irt} + u_{irt}, \tag{5}$$

$$= b_{irt}p_{irt} + \sum_{l=0}^{\mathcal{L}-1} \alpha_l n_{\mathcal{G}_{irt}}^{-1} \sum_{j \in \mathcal{G}_{irt}} e_l(d_{ijrt}^C)p_{jrt} + u_{irt}^*, \tag{6}$$

where d_{ijrt}^C contains the continuous distance measures, $\mathcal{G}_{irt}$ is the set of rival brands of brand i in region r at time t, $n_{\mathcal{G}_{irt}}$ is the number of such brands, the e_l's are the series expansion functionals and the α_l's the corresponding coefficients. u_{irt}^* includes both u_{irt} and the portion of the expansion which is not estimated. What is left is a model linear in coefficients with some endogenous regressors.

Pinkse, Slade and Brett (1998) have shown that both the coefficient estimators of the remaining regressors and the estimator of g have limiting normal distributions. The variance/covariance matrix is consistently estimable even though errors can be correlated across time and space. The procedure suggested by Pinkse, Slade and Brett (1998) is similar to the White (1980) or Newey–West (1987) covariance estimators. Instead of only putting squared residuals on the diagonal of the variance matrix, as in White (1980), we also include cross products of residuals in some of the off–diagonal elements. The cross products correspond to the observations that we expect to have the greatest residual correlation. For consistency, the number of cross products thus included should increase with the sample size, albeit at a slower rate.

Since the expansion terms are interacted with rival prices, the number of endogenous regressors increases with the sample size. Additional instruments can be generated in the same way as if the model were fully parametric: expansion terms can be interacted with rival prices in the other region and/or with exogenous rival characteristics.

2.2 Some Identification Issues

The issue of identification is somewhat different from that in Pinkse, Slade and Brett (1998). First, the expansion is here in terms of variables that can also enter the X–part of the regression model. Second, we now have the potential for an infinite number of rivals (in the limit) at a finite distance from each brand. For example, alcohol contents are unlikely to fall outside of the range 2% to 6%, whereas geographic distance support was assumed to increase with the sample size in our earlier work.[6]

If the number of rival brands at a finite distance increases to a finite limit, then identification–related issues are addressed in our earlier paper. If the number of rival brands at a finite distance increases without bound with the sample size, then the influence of each individual brand is assumed to be asymptotically negligible. In essence, one is then estimating a conditional mean, where the conditioning variable is the brand's location in taste space, say its alcohol level and perhaps the product group to which it belongs.

Since alcohol can also enter as an X–regressor it is not immediately obvious that g is identified, even if alcohol only enters X linearly. However, provided that price distributions are different across regions and/or time periods and if g is the same across regions and time periods, then g can still be identified. For instance, a regression with dependent variable $q_{i2t} - q_{i1t}$ eliminates the alcohol–component of X as well as any time dummies, but not g. Combinations are also possible to simultaneously eliminate time and region fixed effects.

The above procedure is not necessary if discrete distance measures (such as product groupings) are used in g, but no corresponding product dummies are included in X, which is the case in the results that we present in section 7. In general, the procedure will not work well if price distributions and/or locations in taste space do not vary much across categories.

3 Calculation of Endogenous Market Boundaries

The procedure for determining endogenous market boundaries involves a grid search over possible values of the utility–loss parameters b_a, b_c. For each set of values (b_a, b_c) market areas are computed. The size of the market area (in $[0,1] \times [0,1]$) of each brand is then matched up to the corresponding fraction of the sales volume. The market areas corresponding to the value of (b_a, b_c) that minimizes an absolute loss

[6] Commonly used terminology is 'fill–in asymptotics' versus 'increasing domain asymptotics'.

function are then used to determine common boundary relationships.

Below we describe the procedure for finding market areas for specific (b_a, b_c). The procedure consists of three steps detailed in appendices C.1 through C.3, respectively. In the first step, market areas are computed for each observation without regard for the fact that the entire market is contained in $[0,1] \times [0,1]$. In the second step, market areas are 'cropped' to lie inside $[0,1] \times [0,1]$. In the third and final step, the sizes of the market areas are computed.

3.1 Determining Market Areas

We first ignore the fact that taste space is $[0,1] \times [0,1]$ and determine market areas that possibly extend infinitely far. Each market area is a convex polygon (possibly extending to infinity) and contains the set of points at which the utility loss of brand i is less than that of all other brands.

We carry out the following procedure for each brand i.

1. Determine any brands $j \neq i$ with which i shares a boundary extending to infinity. Brands i and j share a common boundary if the set of points at which the utility loss of both brands is the same and is no greater than that of any other brand. The common boundary extends to infinity if the afore–mentioned utility loss can be made arbitrarily large. The number of brands $j \neq i$ with which brand i shares an infinite border is either zero or two.

2. If there are no brands $j \neq i$, for which i and j have common boundaries that extend to infinity, then find a 'corner point' of the market area of i. A corner point is a point at which the utility loss of distinct brands i, j, t is the same and is no greater than that of any other brand. If brand i has no corner points, the imputed size of its market area for the utility–loss parameters used is zero, and the procedure can go on to the next brand.

3. Now find the remaining corner points to complete the convex polygon. There is one other corner point involving brands i, j (and not involving t in case i and j do not share a boundary of infinite length). Find this corner point, which is the next corner of the polygon and involves brands i, j, s, $s \neq i, j, t$. The corner point after the (i, j, s) one involves brands i, s and a third brand, which needs to be determined. Repeat the procedure until the polygon is complete or until the other infinite boundary is reached.

3.2 Cropping

The cropping procedure is used to eliminate all parts of the uncropped market areas that extend outside $[0,1] \times [0,1]$. A full description of the cropping procedure would be too lengthy to include in this paper, but is available from the authors upon request. All brands whose market areas computed in section C.1 are completely outside the box are added to the list of brands with zero market area. Note that brands are not necessarily located inside their own market area.

3.3 Computing Market Area Sizes

We compute areas by noting that a convex polygon is made up of triangles. The area of a triangle is easily computed. Assume that ν_1, ν_2, ν_3 are corner points of a triangle, then $\nu_4 = (1-\alpha)\nu_1 + \alpha\nu_2$, with $\alpha = (\nu_3 - \nu_1)^T(\nu_2 - \nu_1)/||\nu_2 - \nu_1||^2$, is the orthogonal projection of ν_3 onto the line segment connecting ν_1 and ν_2. Then the area of the triangle is $||\nu_1 - \nu_2||\ ||\nu_3 - \nu_4||/2$. We have carved the polygon into triangles as follows. Pick a corner point ν_1^* and move along the polygon labeling the remaining corner points in order $\nu_2^*, \ldots, \nu_{n_\nu}^*$ with n_ν the number of corner points of the polygon. The triangles are those with corner points $\nu_1^*, \nu_2^*, \nu_3^*$, $\nu_1^*, \nu_3^*, \nu_4^*$, $\nu_1^*, \nu_4^*, \nu_5^*$, etcetera.