Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects

Daniel A. Ackerberg∗

and

Marc Rysman∗∗

Commonly used discrete-choice models such as logit, nested logit, and random-coefficients models place very strong restrictions on how unobservable characteristic space changes with the number of products. We argue (and show with Monte Carlo experiments) that these restrictions can lead to biased estimates of price elasticities and the welfare consequences from additional products. In addition, these restrictions can identify parameters that are not intuitively identified given the data at hand. We suggest an alternative model that does not have these properties and present a structural interpretation of the model. Monte Carlo experiments and an empirical example show that this issue can be important in practice.

1. Introduction

The recent literature in applied economics, and empirical Industrial Organization in particular, has often turned to discrete-choice models to estimate demand for differentiated products or different alternatives. In these models, consumer utility functions, market shares, and substitution patterns depend on product characteristics that are observed by the econometrician. In addition, these models typically allow for unobserved product characteristics through the inclusion of some form of “symmetric unobserved product differentiation” (SUPD).1

The most common examples of SUPD are logit errors in consumers’ utility functions (see McFadden, 1974). The economic justification for including unobservable product differentiation is that an econometrician typically does not observe all of the product characteristics that are relevant to consumers’ choices. From an econometric standpoint, allowing for unobservable product

∗ University of California at Los Angeles; ackerber@econ.ucla.edu.
∗∗ Boston University; mrysman@bu.edu.

The authors wish to thank seminar audiences at the Federal Reserve Bank, Board of Governors, Stanford Graduate School of Business, UCLA, Carnegie Mellon, UBC, Johns Hopkins, Georgetown, SITE, the Department of Justice, and the Econometric Society meetings. We also wish to thank Aviv Nevo, Ariel Pakes, two referees, and the Editor for their comments. Particular thanks are due to Anne Hall for her important comments on a later draft of the article. Rysman received financial support from NSF grant no. SES-0112527.

differentiation often prevents these models from predicting zero market shares. Its inclusion can also ease estimation.

We argue that while SUPD in itself may be helpful, commonly used models (e.g., logit models, probit models, nested logit models, and the random-coefficient models of Berry, Levinsohn, and Pakes (1995; henceforth BLP)) implement it in an undesirable way. These models assume that each product added to the market adds one additional dimension to SUPD space. This feature results in very little “congestion” in unobserved characteristic space and can be problematic in situations where different consumers face different numbers of products, because consumers are drawn either from different geographies or from different time periods. We intuitively think that in markets with more products, unobservable characteristic space should “fill up” in some sense. These standard models place strong restrictions on how this occurs.

We show that these restrictions play a major role in econometric identification of two of the major quantities of interest in differentiated-product markets. First are the welfare effects of new products. This problem is one that has been recognized, e.g., in Trajtenberg (1990), Petrin (2002), Berry and Pakes (1999), and Bajari and Benkard (2001). Because of the lack of crowding in the standard treatment of SUPD, welfare calculations in standard models tend to overpredict gains from the introduction of new products. This problem has potentially serious implications for policy issues such as the construction of price indices.

Second and less recognized are the implications of SUPD on estimated substitution patterns. We argue that using the standard versions of SUPD can lead to misleading econometric conclusions regarding price elasticities, in terms of both magnitudes and statistical significance. Restrictions of standard SUPD force variation in the number of products in the choice set to identify (or help identify) price elasticities. Interestingly, we show that with these restrictions, one can often “identify” price elasticities without observing meaningful variation in prices. This source of identification relies entirely on assumptions about unobservable characteristic space. These assumptions are even more unreliable if, as is often the case, “defining” different products has some arbitrariness to it.

There are two previous approaches in the literature that address these issues. The first set of work (e.g., BLP (1995) and Petrin (2002)) tries to reduce the importance of SUPD by linking substitution patterns to observable continuous characteristics (e.g., BLP) or observed groupings (e.g., the nested logit). This approach keeps SUPD (e.g., logit errors) in the model but attempts to reduce its importance. These methodologies work to the extent that the econometrician observes the relevant product characteristics. However, as inflexible SUPD still exists in these models, its effects can still exist.

A second and more recent approach, advocated by Berry and Pakes (1999) and Bajari and Benkard (2002), eliminates SUPD altogether from the model. In their “pure hedonic” models, products are unobservably differentiated only with respect to a single dimensional unobserved characteristic. As new products enter, this unobserved characteristic space becomes more crowded.


3 While we believe that our methodologies will provide “better” estimates of these welfare effects, the welfare gains of any new product will depend on the shape of the demand curve at very high prices. Thus, unless one observes a wide range of prices, any calculation of welfare gains is going to rely on fairly “structural” assumptions about the upper portion of the demand curve.

4 For example, with cars and computers, the empirical definition of what constitutes a “choice” clearly has some arbitrariness to it (e.g., BMW 3 Series versus (BMW 330, BMW 325) versus (BMW 330i, BMW 330Ci, BMW 330 Ci convertible)).

5 In addition to logit errors, these approaches typically allow for a scalar unobserved product characteristic corresponding to each product. However, as these scalar unobserved product characteristics are equally valued by all consumers, they do not play a major role in determining the extent of product differentiation.

6 Feenstra and Levinsohn (1995) also estimate a multidimensional pure hedonic model, albeit without any unobserved characteristics.
While these approaches are intuitively very attractive in the sense that there are no ad hoc logit errors, they are also either more computationally intensive (Berry and Pakes) or more data intensive (Bajari and Benkard) than standard models including a logit error.\(^7\)

This article suggests a third approach, which we interpret as somewhat of a compromise between the above two. We argue that it is the unnecessary inflexibility of standard logit errors that can adversely affect estimates of parameters of interest such as substitution patterns and welfare effects. As such, we keep logit errors in our model but allow them to be considerably more flexible than is currently done. This flexibility allows an econometrician to estimate how fast unobserved characteristic space expands with the addition of new products, not assume it as prior work does. In practice, our approach simply puts functions of the number of products in a market (and/or the number of products in a group or nest) into the discrete-choice estimating equation. We show that this model has a structural interpretation—one where new products crowd out existing products in retail store or shelf space. Although this model of “crowding out” is very stylized, it is intuitive and captures phenomena that we believe actually occur in markets.\(^8\) Our flexible logit error imposes no additional computational burden, and thus our approach is considerably simpler to implement than Berry and Pakes (1999) as well as less data intensive than Bajari and Benkard (2002). On the other hand, those with a more structural leaning might prefer their methods in that they completely eliminate ad hoc logit errors, while we only make them more flexible.\(^9\)

We proceed as follows. In Section 2 we argue (1) that traditional discrete-choice models place unnecessary restrictions on \(SUPD\), (2) that these restrictions can “identify” parameters that intuitively should not be identified, and (3) that these restrictions can bias estimates of parameters of interest. Section 3 introduces our model of product congestion and discusses estimation. In Section 4 we present Monte Carlo results showing that in the presence of product congestion, standard estimation procedures can give biased results (sometimes very large) and that these biases tend to be in particular directions. Section 5 applies the estimator to data on Yellow Pages demand from Rysman (2004). We find that the adjustments significantly affect predictions. Section 6 discusses a multiplicative adjustment, which provides many of the same benefits for estimation with a slightly different theoretical motivation.

Lastly, note that much of our applications are focused on the context of estimating aggregated discrete-choice models. These models are typically estimated on data across markets (in space or time) where one often observes changes in the size of the choice set and where our concerns are relevant. However, our comments and techniques are equally applicable for discrete-choice models estimated on individual-level data (e.g., product, employment, or transportation choice) when there are changes in the choice set over individuals or time.

### 2. Unobserved differentiation in common discrete-choice models

This section argues that assumptions about unobservable characteristic space used in traditional discrete-choice models are restrictive, and that this leads to undesirable identification results. We briefly suggest our solution to the problem, which is formalized and further motivated in Section 3. We use the nested logit model to formally illustrate our identification points, but we discuss the extension of our arguments to a full random-coefficients model.

\(^7\) Another possible critique of these hedonic models is that while unobserved characteristic space may expand too much with logit errors, it may expand too little with the pure hedonic models, at least when unobserved characteristic space is modelled as one-dimensional.

\(^8\) For example, retail stores often sell only a small subset of the available wholesale products. Computer retailers, e.g., typically display between 10 and 30 computers, while the total number of wholesale computers available in a given year is between 150 and 250 (Pakes, 2003). Presumably, this is due to the costs of retail space.

\(^9\) An interesting issue is to what extent these various models are observationally equivalent in terms of market shares, market share derivatives, or market share changes with new products (see, e.g., Anderson, DePalma, and Thisse (1992) and McFadden and Train (2000)). Regardless of observational equivalence, the fact that the literature has gravitated toward using models including logit errors makes our results relevant.
Identification. We start by using derivative-based identification arguments to show how the nested logit model handles economically interesting variation in a restrictive way. For exposition, assume there are $J$ products and an outside option, labelled product 0. The $J$ products are in one group (nest) $g$ and the outside option is in a group by itself. In the nested logit model, the utility obtained by consumer $i$ from product $j$ ($j > 0$) is

$$u_{ij} = \beta_0 + X_j \beta_1 + \zeta_{ig}(\sigma) + \varepsilon_{ij},$$

where $\varepsilon_{ij}$ is distributed type-I extreme value, and $\zeta_{ig}(\sigma)$ is constant for each individual across the product nest and distributed such that $\zeta_{ig}(\sigma) + \varepsilon_{ij}$ is distributed type-I extreme value (see Cardell, 1997). Note that $\varepsilon_{ij}$ represents consumer $i$’s idiosyncratic taste for good $j$ and $\zeta_{ig}$ represents $i$’s idiosyncratic taste for products in group $g$. As is standard, we assume $u_{i0} = \zeta_{i0}(\sigma) + \varepsilon_{i0}$, normalizing the “mean” utility of the outside option to zero. The parameter $\sigma \in [0, 1]$ measures correlation in unobserved utility among products in the nest. Lower values of $\sigma$ imply stronger within-group substitution relative to across-group substitution (in this case, substitution to the outside alternative). In what follows, we interpret $X_j$ as the price of product $j$, but our arguments trivially apply to elasticities with respect to general product characteristics.

Denote the market share for firm $j$ as $s_j$, the market share for the entire group of inside products as $s_g$, and the market share for $j$ within group $g$ as $s_{j|g}$. We then have (Cardell, 1997)

$$s_{j|g} = \frac{\exp \left( \beta_0 + X_j \beta_1 \right)}{\sum_{k=1}^{J} \exp \left( \beta_0 + X_k \beta_1 \right)},$$

$$s_g = \frac{D\sigma}{1 + D\sigma},$$

$$s_j = s_{j|g} s_g,$$  

(1)

where $D = \sum_{k=1}^{J} \exp \left( \beta_0 + X_k \beta_1 \right)$.

There are three forms of variation in data that identify the parameters $\beta_1$ and $\sigma$ in the nested logit model. The first type is variation in within-group market shares due to changes in observable product characteristics. Looking at the derivative corresponding to this type of variation tells us what parameters are identified by the variation. The derivative is

$$\frac{\partial s_{j|g}}{\partial X_j} = \beta_1 s_{j|g} (1 - s_{j|g}),$$

(2)

suggesting that this type of variation identifies $\beta_1$.

The second and third types of variation are changes in group market shares ($s_g$) due to (1) changes in observable product characteristics and (2) changes in the number of products. To focus on group-level changes, assume $X_j = X \forall j$. In that case, the derivatives of group share $s_g$ with respect to $X$ and $J$ are

$$\frac{\partial s_g}{\partial X} = \sigma \beta_1 s_g (1 - s_g),$$

$$\frac{\partial s_g}{\partial J} = \sigma \frac{s_g (1 - s_g)}{J}.$$

(3)

This suggests that there are two sources of identification for $\sigma$: cross-group switching from changes in the number of products and cross-group switching from changes in observed characteristics. Note that there are also two sources of identification for $\beta_1$: within-group switching from changes

10 We ignore endogeneity issues regarding price, which has been a focus of the prior literature. The points in our article are valid whether price movements are purely exogenous or whether they are endogenous and one must find some exogenous source of price variation.

11 Note that the normalization (and notation) used above and in the following identification arguments is different than the normalization used by Berry (1994).

12 Note that the constant term $\beta_0$ is identified by the level of the inside product market shares.

13 Note that these comparative statics correspond to hypothetical “experiments” we would like to do in the data to identify parameters.
in observed characteristics and cross-group switching from changes in observed characteristics. Given that these three comparative statics ($\partial s_g/\partial X$, $\partial s_g/\partial J$, and $\partial s_j/\partial X_j$) map into only two structural parameters ($\beta_1$ and $\sigma$), the model implies a restrictive relationship between the effects.

This restrictive relationship has interesting implications for identification. Observing markets where product characteristics (or price) differ across markets but the number of products is the same in all markets can identify both $\sigma$ and $\beta_1$. Therefore, a researcher can identify the effects of adding a product to the choice set (e.g., the additional welfare generated by the new product) without ever observing variation in the number of products. Perhaps even more unintuitively, one can identify cross-price elasticities between products in the group without ever observing changes in relative prices of the products (for a simple example of this, see the first part of Section 3). More generally, not only do price changes play a role in identifying price elasticities, but less intuitively, changes in the number of products play a role in identifying price elasticities. Similarly, not only do observed changes in the number of products play a role in identifying the effects of changing the number of products, less intuitively, changes in prices or characteristics will play a role in identifying these effects.14

Are these unintuitive sources of identification believable? Clearly this identification is coming from something in the structure of the demand model. Thus, the answer to this question should depend on whether this structure is believable. We argue through the rest of the article that what is generating this identification is a very peculiar and unintuitive property of standard logit errors. As such, our answer to the above question would be “no.”

\section{Properties of logit errors.}

Any model including logit errors implicitly makes restrictive assumptions about the relationship between unobservable characteristic space and the number of products. Specifically, logit errors imply that the dimension of unobservable characteristic space expands proportionally to the number of products. To see this, note that we can write consumer $i$’s set of logit errors for the $J$ products as

\[
\begin{align*}
\epsilon_{i1} &= d_{11}\epsilon_{i1} + \cdots + d_{1J}\epsilon_{iJ} \\
& \vdots \\
\epsilon_{iJ} &= d_{J1}\epsilon_{i1} + \cdots + d_{JJ}\epsilon_{iJ},
\end{align*}
\]

where $d_{jk}$ are dummy variables with $d_{jk} = 1$ if and only if $j = k$, $d_{jk} = 0$ otherwise. Written in this way, we can interpret logit errors as representing a $J$-dimensional characteristic space: $(\epsilon_{i1}, \ldots, \epsilon_{iJ})$ are consumer $i$’s preferences over the $J$ dimensions, and the vector $(d_{j1}, \ldots, d_{JJ})$ represents product $j$’s “location” in the $J$-dimensional space.

With this interpretation, note that if we add another product $(J+1)$ to the model, this product differentiates in an entirely new dimension (that of $d_{J+1, J+1}$), which is associated with a new logit error $\epsilon_{iJ+1}$. Thus, the dimensionality of unobserved characteristic space expands by 1 with the addition of the new product.

Another implication of logit errors is that all products are “equidistant” from each other in unobserved characteristic space and this distance remains constant as the number of products in the market changes. In a sense, there is no “crowding out” or “congestion” in unobserved characteristic space. This is counterintuitive in the following way. With classical product-differentiation models

\footnote{One’s choice of instruments can affect how these comparative statics play a role in identifying parameters. For example, suppose one has the choice of using $(1/J) \sum x_j$ and/or $J$ as an instrument (for within-group share in the Berry (1994) inversion) in the nested logit model. Using only $(1/J) \sum x_j$ as an instrument would correspond to fitting the second (third) comparative static more closely and would probably lead to better (worse) estimates of price elasticities and worse (better) estimates of welfare effects. If one uses both instruments (or a combination of the two, e.g., $\sum x_j$ as suggested in BLP), which comparative static is fitted “better” will depend on the relative amounts of variation in $(1/J) \sum x_j$ and $J$ in the data.}
such as the Hotelling model or the Salop circular model in mind, one would naturally expect products in more dense markets to be “closer” in characteristic space.15

As will become clear later, it is these strong assumptions about the relationship between unobservable characteristic space and the number of products that generate the unintuitive identification results above. Therefore, unless one completely believes in this “no crowding out” property of logit errors, one should probably not believe these sources of identification and worry about obtaining biased estimates of parameters (e.g., $\sigma$ and $\beta_1$), price elasticities, and welfare calculations.16

□ More-general models. The arguments of the first part of this section were based on a fairly simple nested logit model. Do random-coefficients models with logit errors (e.g., BLP, McFadden and Train (2000), and Nevo (2001)) or nested logit models with more-complicated nesting structures have similar identification properties? We believe so. Comparative statics in these models are too complicated to formulate arguments like the above. However, we can appeal to a number of informal arguments. First, note that the nested logit model is in fact a random-coefficients model—one where the random coefficient is on a group dummy variable and has a particular distribution (parameterized by $\sigma$). The intuition behind these identification results should not change if random coefficients are instead on continuous characteristics and/or are assumed to have normal distributions.17 Second, as with the nested logit model, all of the parameters would be identified if one estimated a random-coefficients model on a set of markets all with the same number of products. Therefore, any variation due to the fact that markets have different numbers of products is necessarily handled in a restrictive way. Third, our Monte Carlo results on random-coefficient models suggest that they have similar problems.

Generally, we believe that any model including standard logit errors will have similar properties, and that identification in these models is suspect unless one believes the unintuitive and restrictive assumptions inherent in standard logit errors.

□ Proposed solution. We now briefly preview our proposed solution to the problem, showing that it eliminates the perverse identification results discussed above. Later, in Section 3, we give a structural interpretation of our solution. This structural interpretation corresponds to relaxing the “no crowding out” assumption of standard logit errors.

We propose adding a function $f(J; \gamma)$ with parameter $\gamma$ to the term $\beta_0 + X_j \beta_1$ in the nested logit model (1), i.e.,

$$s_{j|g} = \frac{\exp(\beta_0 + X_j \beta_1 + f(J; \gamma))}{\sum_{k=1}^{J} \exp(\beta_0 + X_k \beta_1 + f(J; \gamma))},$$

$$s_g = \frac{D^g}{1 + D^g},$$

$$s_j = s_{j|g} s_g,$$ (4)

where $D = \sum_{k=1}^{J} \exp(\beta_0 + X_k \beta_1 + f(J; \gamma))$. With this model, the three comparative statics discussed above are

$$\frac{\partial s_{j|g}}{\partial X_j} = \beta_1 s_{j|g} (1 - s_{j|g}),$$

$$\frac{\partial s_g}{\partial X} = \sigma \beta_1 s_g (1 - s_g),$$

$$\frac{\partial s_g}{\partial J} = \sigma s_g (1 - s_g) \left( \frac{1}{J} + f^*(J; \gamma) \right).$$ (5)

15 In logit characteristic space, if one randomly chooses two products from a market, the expected difference between $s_{j_1}$ and $s_{j_2}$ is the same regardless of the number of products in the markets. In contrast, consider a Hotelling model where products space themselves out as much as possible. With two products in the market, the expected distance between two randomly chosen products (without replacement) is trivially 1, with three products in the market the expected distance is $1/3 + 1/3 + 1/3 = 1/3$, with four products it is $1/4 + 1/4 + 1/4 + 1/4 = 1/2$, and with five products it is $1/5 + 1/5 + 1/5 + 1/5 + 1/5 = 2/5$.

16 The CES demand system also does not display crowding, and is in fact subject to many of the criticisms about elasticities and welfare effects that we make of logit-based models. Extensions of our adjustment to the CES model are available from the authors.

17 In other words, we expect that in these models, comparative statics in entry and exit will also play a role in identifying price elasticities, and comparative statics in prices will play a role in identifying the effects of additional products. Again, this identification is likely to be highly reliant on the exact structure of errors.
The first two comparative statics are the same as before, but the third now depends on an additional parameter, $\gamma$, in the new function.\(^{18}\)

Note that this adjustment gives the nested logit model the ability to match all of the observed variation in the data. In this model, $\beta_1$ will be identified by variation in within-group market shares due to changes in observable product characteristics. Conditional on $\beta$, $\sigma$ is identified by changes in group market share in response to changes in observable product characteristics, and conditional on $\sigma$, $\gamma$ is identified by changes in group market share in response to changes in the number of products. In this adjusted model, one cannot identify the effect of adding a product to the choice set without observing variation in the number of products, nor identify cross-price elasticities between products in the group without observing changes in relative prices of the products. More generally, in this model we expect price elasticities to be identified by price variation, not by changes in the number of products. Similarly, effects of changing the number of products should be identified by actual changes in the number of products, not by changes in prices. In essence, this adjusted model eliminates the unintuitive sources of identification described earlier.

3. A structural interpretation

In this section we exhibit a structural model that generates the adjustment suggested in the previous section. This provides a structural interpretation of the new parameters, which can aid in understanding and adding further to the model (for instance, writing a first-order condition for the producers). It also shows how the adjustment detailed above corresponds to a more flexible logit error that eliminates the “no crowding out” assumption of standard logit errors. We also discuss estimation issues.

Intuition behind the model. We begin with a story. Suppose one is interested in estimating a nested logit model of competition between fast food firms (one nest is the fast food restaurants and one nest is a composite “outside” good). Data is obtained on prices and market shares for two time periods of data. In the first time period there is only one firm, $MD$, and in the second period there is entry and thus two firms, $MD$ and $BK$. Suppose that prices are identical for all firms in all periods, that in the first period $MD$ has a 50% market share, and that in the second period both $MD$ and $BK$ have 25% market shares.

Since the entry of $BK$ “steals” market share only from $MD$ (and not the outside alternative), a nested logit model will necessarily estimate $\sigma = 0$, i.e., that the within-group variance is zero. This $\sigma = 0$ implies (1) that $MD$ and $BK$ are identical in all respects to all consumers, and (2) that the cross-price elasticity between $MD$ and $BK$ is infinite. Note that identification here has come solely from changes in the number of products, as there is no variation in prices.

Now consider an alternative story of what is going on in this data. Suppose these firms operate through outlets (franchises) and there is important geographical differentiation (i.e., all else equal, consumers tend to go to the nearest outlet location). Other than geographic differentiation through their outlet locations, the food served by $BK$ and $MD$ is identical. In the first period there are two outlets, both franchised to $MD$. In the second period there are also two outlets, but one of the $MD$ outlets has been taken over by $BK$. Since prices remain constant and $MD$ and $BK$ serve identical food, this story is perfectly consistent with the market share data above. But is the nested logit prediction of infinite price elasticities correct in this example? We would expect not. Due to the geographic differentiation, we would expect a price cut by $BK$ to only partially cut into $MD$’s market share. The nested logit model estimate of $\sigma = 0$ is highly misleading here: unintuitive restrictions of the model (rather than valid price variation) are incorrectly identifying price elasticities to be infinite.

The intuition behind this story can motivate a structural model in which $J$ enters the discrete-choice estimating equation. In the example, unobserved characteristic space (in this case, outlet

\(^{18}\) Note that our adjusted model is somewhat in the spirit of McFadden’s (1975) “universal logit” model, which somewhat arbitrarily includes characteristics of all products in the utility function for a particular product. In contrast, we focus on a specific adjustment and provide a structural model generating this adjustment.
locations) is subject to congestion: the entry by $BK$ reduces the number of outlets $MD$ has. This “crowding” at the outlet level confounds the observation that a new product has entered. Standard logit-based models simply do not deal well with such congestion, hence the incorrectly predicted price elasticities. We now present a formal model of such retail crowding or product congestion that deals with this issue. If we were to take this model to the fast food data described above, price elasticities would not be identified—an intuitive outcome given the lack of any variation in prices.

□ A model of product congestion. Suppose that the products of interest are sold through a retail market consisting of $R$ retail outlets. As in the above example, we consider the standard case where market shares are observed at the product level: data at the retail outlet level are not observed. Modelling unobserved retail outlets is simply a way of motivating our more general logit errors. Assume that each retail outlet sells only one of the wholesale products, and that product $j$ is sold in $R_j$ retail outlets where $\sum_j R_j = R$. The twist of our congestion model is that logit errors represent idiosyncratic, unobserved consumer preferences over retail outlets rather than over products. (In the next section we expand the model to one in which consumers have logit errors based around both retail outlets and products.) Precisely, the logit utility function for consumer $i$ purchasing from retail outlet $r$ takes the form

$$U_{ijr} = u_j + \varepsilon_{ir},$$

where $u_j$ measures mean product quality. A typical specification for $u_j$ is $u_j = X_j \beta - \alpha p_j + \xi_j$, where $(X_j, \xi_j)$ are product $j$’s characteristics (observed and unobserved respectively) and $p_j$ is its price. The important distinction between this and a standard logit model is that it contains $\varepsilon_{ir}$, not $\varepsilon_{ij}$. Intuitively, $\varepsilon_{ir}$ might capture the fact that consumers live different distances from the $R$ retail outlets.

Note how this model captures congestion as new products enter the market. In the standard logit model, when new products enter the market, new $\varepsilon_{ij}$ are drawn for the new products. In the extreme version of our congestion model, where the number of retail stores $R$ does not change as new products enter, there are no new unobservable terms drawn. The dimensionality of the unobserved characteristic space remains the same as the new products simply crowd out the old products from retail stores.

To aggregate the model to the level of observation (the product level), we need to aggregate over retail outlets. The share of product $j$ is the sum of the shares of all the retail outlets that carry product $j$. As the probability that $i$ buys from $r$ is the same across outlets that carry $j$, the market share for product $j$ is

$$s_j = \frac{R_j e^{u_j}}{1 + \sum_k R_k e^{u_k}} \quad \text{(6)}$$

$$= \frac{e^{u_j + \ln(R_j)}}{1 + \sum_k e^{u_k + \ln(R_k)}}. \quad \text{(7)}$$

Note that the difference between our congestion logit model and a standard logit model is simply the additional term $\ln(R_j)$ in the market share function.

□ Estimating the model. With individual-level data, (6) could be estimated by maximum likelihood. With aggregate data, this model can be estimated using the Berry (1994) inversion:

$$\ln \left( \frac{s_j}{s_0} \right) = u_j + \ln(R_j).$$

In practice, one needs to parametrically specify $R_j$. In the simplest case, where each product is sold in an equal number of retail stores, we have $R_j = \frac{R}{j}$ and we need only specify $R$. One
example is

\[ R = \gamma_0 + \gamma_1 J, \]

where \( J \) is the number of products. As scaling up \( R \) is unidentifiable from the constant term in the utility function, a normalization is necessary, an obvious one being

\[ R = \gamma + (1 - \gamma) J. \]

This results in the estimating equation

\[ \ln \left( \frac{s_j}{s_0} \right) = u_j + \ln(\gamma / J + 1 - \gamma). \] (8)

This specification is attractive in that it nests the pure logit model (\( \gamma = 0 \)) as well as the pure congestion model (\( \gamma = 1 \)). With \( \gamma = 0 \), the number of retail outlets (and correspondingly the dimension of SUPD) increases proportionally to the number of products, whereas with \( \gamma = 1 \) it does not change in the number of products. Intermediate cases are captured by \( 0 < \gamma < 1 \).

Another suggestion for parameterizing the additive term is to let \( \ln(R_j) = \gamma \ln(J) \). In this case, \( \gamma = 0 \) is still the standard logit model and \( \gamma = -1 \) is still a full crowding model (in the sense that expected welfare depends on observable product characteristics but not the number of products). A nice attribute of this specification is that in contrast to (8), this specification can be estimated with OLS or IV techniques. A drawback is that this specification lacks a clear structural interpretation of the parameter. Last, note that one might estimate \( R(J) \) nonparametrically. Given that \( J \) is discrete, this is extremely simple: one just includes indicator functions for different market size (with a normalization for one \( J \)).

**Extensions of the model.** The assumption that all products are sold by an equal number of retail stores might not seem reasonable. However, given no data on retailers, it is hard to imagine how one could intuitively separate out effects of product characteristics and price on utilities versus their effects on the number of retail stores carrying the product. To formalize this, suppose that

\[ R_j = f(J) e^{X_{ij} \beta_1}, \]

so that product characteristics do affect \( R_j \). In this case, \( \tau_1 \) is not separately identified from \( \beta \), the parameters in the utility function. With other specifications of \( R_j \), the different effects might be identified computationally, but this identification would be completely dependent on nonlinearities. As such, we suggest the specification where all products are sold by an equal number of stores.

The assumption that logit errors are not correlated for the same product sold across different outlets may also seem unreasonable. However, we can obtain a very similar estimating equation in a model that relaxes this assumption. Suppose consumers have unobserved tastes over both products and retail stores, i.e.,

\[ U_{ijr} = u_j + \epsilon_{ij} + \rho \epsilon_{ijr}. \]

\( \epsilon_{ij} \) is consumer \( i \)'s product-specific taste, \( \epsilon_{ijr} \) is consumer \( i \)'s product retail outlet–specific taste, and \( \rho \) is a weighting parameter that measures the relative importance of the two unobservables. This formulation is very similar to the standard nested logit model. With the standard nested logit distributional assumptions (\( \epsilon_{ijr} \) distributed type-I extreme value, \( \epsilon_{ij} \) distributed such that \( \epsilon_{ij} + \rho \epsilon_{ijr} \) distributed type-I extreme value), we get the following product-level market shares:

\[
 s_j = \frac{\left[ R_j \exp \left( \frac{u_j}{\rho} \right) \right]^\rho}{1 + \sum_k \left[ R_k \exp \left( \frac{u_k}{\rho} \right) \right]^\rho} = \frac{\exp(u_j + \rho \ln(R_j))}{1 + \sum_k \exp(u_k + \rho \ln(R_k))},
\]
where $R_j$ is the number of retail stores in which product $j$ is sold. Then we have the estimating equation,

$$\ln \left( \frac{s_j}{s_0} \right) = u_j + \rho \ln(R_j).$$

Consider the specification $\ln(R_j) = \gamma \ln(J)$. In this case, $\gamma$ and $\rho$ are not separately identified, only their product $\rho \gamma$ is. While a different specification for $R_j$ might lead to separate identification of $\rho$ and $\gamma$, it would again be based on a nonlinearity. This lack of identification is not a drawback, because separating the parameters (e.g., $\rho$ versus $\gamma$) is irrelevant for empirical or welfare implications. It means that our original model is robust to unobserved tastes at both the product and retail store level.

□ Application to more-general discrete-choice models. The above subsections added congestion to a simple logit model. We can similarly add congestion to more-realistic models such as nested logit and random-coefficients models. For example, consider the nested logit utility function:

$$U_{ijr} = u_j + \zeta_{ig} + \epsilon_{ir},$$

where $\zeta_{ig}$ is consumer $i$’s idiosyncratic tastes for products in group $g$. Note that this nested error term is defined over product groupings and not retail store groupings (since retail stores are not observable, one cannot group them). The variable $\epsilon_{ir}$ is still a retail store–specific unobservable. With this utility function, product shares are given by

$$s_j = s_{j|g} s_{g} = \left[ \frac{R_j \epsilon_{j\gamma}^{\frac{\sigma}{\gamma}}} {\sum_{k \in g \neq j} R_k \epsilon_{k\gamma}^{\frac{\sigma}{\gamma}}} \right] \left[ \frac{\left( \sum_{k \in g \neq j} R_k \epsilon_{k\gamma}^{\frac{\sigma}{\gamma}} \right)^{\gamma}} {1 + \sum_{g} (\sum_{k \in g \neq j} R_k \epsilon_{k\gamma}^{\frac{\sigma}{\gamma}})^{\gamma}} \right],$$

and estimation can proceed using the Berry (1994) inversion:

$$\ln \left( \frac{s_j}{s_0} \right) = u_j + \sigma \ln(R_j) + (1 - \sigma) \ln s_{j|g}.$$  

Again, we need to parameterize $\ln(R_j)$ to estimate this model. The simplest approach would be to do exactly what we did in the logit model, specifying $\ln(R_j)$ as equal to either $\ln(J/J + 1 - \gamma)$ or $\gamma \ln(J)$. As a more ambitious and flexible alternative, one might want to allow congestion to vary across products. In other words, one might expect goods in one nest to crowd out (in terms of retail space) goods in the same nest more than goods in different nests. One could accommodate this possibility by, e.g., allowing $R_j$ to depend on the number of products in the nest as well as the total number of products. With multiple-level nested logit models or other GEV models (e.g., the model of Bresnahan, Stern, and Trajtenberg (1997)), one could allow more complex $R_j$ functions.

In random-coefficients models like BLP, one could again simply add $\ln(J/J + 1 - \gamma)$ or $\gamma \ln(J)$ to the conditional (on random coefficients) market share equations. Again, while this allows congestion, it assumes that the congestion occurs equally across products. A more flexible approach might let $R_j$ be a weighted count of the number of products in the market, where the weights depend on how close other products are to $j$ in characteristic space. For instance, one could specify $R_j$ as

$$R_j = \sum_{k=1}^{J} \phi((X_j - X_k)^*(\text{cov}(X))^{-1}(X_j - X_k)),$$

Note that we now (and in the rest of the article) use Berry’s (1994) normalization, although we still use our redefined $\sigma$. Formally, to transform these parameters (and the parameters in equation (9)) to Berry’s parameters, use $\sigma = 1 - \sigma_{\text{Berry}}$ and $\beta = \beta_{\text{Berry}}$. © RAND 2005.
where $\phi$ is the normal. This specification is similar in spirit to counting products in the same nest differently than products in different nests in the nested logit model. Intuitively, researchers might expect that products that are close together in observed space crowd each other out more than more distant products; this specification allows for this possibility.\footnote{Note that the McFadden and Train (2000) result regarding the generality of the mixed multinomial logit (or random-coefficients) model does not apply to data generated from our model. The reason is that in our congestion model, the distribution of the unobservable term for each wholesale product depends on the number of other wholesale products.}

### 4. Monte Carlo results

In this section we use Monte Carlo simulations to study how standard logit-based models perform when data are generated according to our congestion models. In particular, we examine how well the standard models estimate price elasticities and the welfare effects of new product introductions. We find potentially large biases in both quantities across a variety of specifications, suggesting that ignoring congestion can be problematic in practice.

#### Nested logit model

The rows of Table 1 contain various specifications of our congestion model in a nested logit framework. In all specifications, we simulate data from a very large number of markets ($N = 5,000$). Because of this large amount of data, there is very little estimation error in our estimates (and resulting elasticities), so these estimates can essentially be interpreted as asymptotic results. In each market, there are between 1 and 10 products, distributed uniformly across this range. There are two nests in each market; the first contains all the inside products, the second contains only the outside alternative. In the base specification, price is drawn from a normal distribution with mean 2 and variance 0.2.\footnote{Half of this variation in price is within-market, half is across-market.} The constant term in utility is 1 and the coefficient on price is $-1$. The nested logit parameter $\sigma$ is initially set at 0.8. As is standard, the utility from the outside alternative is normalized to zero.

The various specifications in the rows of the table change various parameters of the model. The “nested logit” subrows contain the results of naive nested logit estimation on these data, using the standard Berry (1994) inversion with the number of products in the market ($J$) and the mean characteristic in the market ($\bar{X}_j = \sum X_j / J$) as instruments for the endogenous $s_{jg}$.\footnote{In generating our data, we also allowed a scalar unobserved characteristic valued equally by consumers, $\xi_j$, to generate an econometric error at the aggregate level (see Berry, 1994). The variance of $\xi_j$ across products was set at 0.5.} Because of the large amount of data, the “truth” subrows in the tables are not only the estimation results from our congestion models, but also the true values of these quantities (since we use a large amount of data and the truth is our congestion model). The columns of the table contain various elasticity and welfare calculations at the estimates. Elasticities are computed for the mean market with $J = 5$ and $X_j = 2 \forall j$. Cross-price elasticity is $\partial s_k / \partial X_j (X_j / s_k)$, outside-good price elasticity is $\partial s_0 / \partial X_j (X_j / s_0)$. Welfare increase refers to the percentage increase in welfare moving from a market with 1 product to a market with 10 products.

The first row of Table 1 contains results for the full congestion model. In this model $\gamma = 1$, i.e., the number of retail outlets does not change as the number of products increases. Naive nested logit estimation of this model gives extremely poor results. The nested logit estimates the average own-price elasticity to be $-11.07$, while the actual own-price elasticity is $-2.29$. Within-group cross-price elasticities are off by an order of magnitude, and estimates of across-group (to the outside alternative) price elasticities are about 70% of their true value. While in actuality there is no welfare gain moving from 1 product to 10 products (since in the full-congestion model new products “completely” crowd out the old ones), the nested logit estimates suggest a gain of 20%. Interestingly, in this case the nested logit model does a reasonable job at matching welfare gains (at least in an absolute sense), but a terrible job at price elasticities.\footnote{This does match the fast food franchise story in Section 3, where the nested logit model predicts $\sigma = 0$, thus correctly measuring the welfare gains due to the entry of BK to be zero.}
There is a clear intuition as to why, in the presence of congestion, standard estimation methods are prone to overestimate within-group cross-price elasticities and underestimate across-group cross-price elasticities. The standard nested logit specification underestimates the nesting parameter $\sigma$ (e.g., in row 1 the standard nested logit model estimates $\sigma = .093$ while in truth, $\sigma = .8$). Consider again the estimating equation under our adjustment:

$$\ln\left(\frac{s_j}{s_0}\right) = X_j\beta - ap_j + (1 - \sigma)\ln(s_{j|R}) + \sigma\ln(R_j(J)) + \xi_j. \quad (9)$$

The standard approach ignores the term $\sigma\ln(R_j(J))$. Recall that $R_j(J)$ will decline in $J$ if there is any congestion, i.e., if the number of retail stores in which product $j$ is sold declines in $J$. Typically the within-group share, $\ln(s_{j|R})$, will also decline in $J$, so the omitted variable will be positively correlated with $\ln(s_{j|R})$ (and one of the instruments for $\ln(s_{j|R})$, $J$). This will tend to bias the estimate of $\sigma$ downward in the standard nested logit model. The underestimate of $\sigma$ suggests too much insulation between groups. As such, across-group substitution is estimated to be too weak, and within-group substitution too strong.

Rows 2 through 9 perturb the parameters of the model. In rows 2 through 4, the congestion parameter $\gamma$ is varied. As would be expected, the nested logit estimates are closer to the truth as $\gamma$ decreases (recall that $\gamma = 0$ implies no congestion, i.e., the standard nested logit model is the truth). However, even at $\gamma = .5$, there are still significant biases in the nested logit results. Row 5 changes the nesting parameter $\sigma$ from .8 to .2. While the nested logit does a bit better on price elasticities (proportionally), it does worse with welfare predictions. Rows 6 through 8 respectively change the constant term in the utility function, the slope term in the utility function, and the mean of price. The large biases in price elasticities and welfare calculations continue to persist. In the last row of the table, the variance of price is increased in the simulated dataset. Interestingly, estimates of price elasticities get considerably better, while estimates of welfare changes worsen. We believe the intuition behind this result is that increasing the variance in price increases the data’s information on the second comparative static (in Section 2) relative to the third comparative static. This will tend to move parameters such that the second comparative static is more closely satisfied, but the third comparative static is less closely satisfied. Since the second comparative static is directly related to elasticities, while the third comparative static is more related to welfare changes due to changes in the number of products, this improves price elasticity estimates, but worsens estimates of welfare effects.

\(\square\) Random-coefficients (BLP) model. We also simulate a random-coefficients logit model. This is the type of model used in BLP. We again use 5,000 markets with the number of products distributed uniformly from 1 to 10. Price is again drawn from a normal distribution with mean 2 and variance .2. The constant term in the utility function is initially set at 2. We allow a random coefficient on price, equal to $\beta_1 = \exp(\sigma_\beta z)$, where $z$ is a standard normal, and initially, $\beta_1 = -2$ and $\sigma_\beta = .2$. We impose that the data is drawn from a crowding model with crowding term $\ln(\gamma/J + 1 - \gamma)$. In Table 2, rows marked “RCM” correspond to estimates of a naive BLP-style random-coefficients model with a regular logit error (using a constant, $X_j$, $J$, and $(1/J)\sum X_j$ as instruments). Rows marked “truth” correspond to estimates of a random-coefficients model with our more flexible logit error (again, this corresponds to the true elasticities and welfare effects).

Examining the table, the first row considers the case where $\gamma = .95$. As in the nested logit case, both elasticities and welfare calculations are considerably biased with the naive RCM. Again, as we lower the level of crowding, the standard RCM does better, but there are still significant biases when $\gamma = .5$. The remaining rows in the table again perturb the parameters of the model, and again not much changes: biases in both price elasticity and welfare calculations persist. As in the nested logit results, increasing the variance of the observed characteristic in the last row improves price elasticities but worsens welfare calculations. The worsening of welfare calculations is marginal, though. This may be due to the fact that as the variance of the observable characteristic increases, the relative importance of logit errors in the model decreases. One would expect this effect to tend
to improve both price elasticities and welfare calculations, perhaps counteracting the worsening of welfare calculations due to the comparative static effect.\textsuperscript{24} In summary, our Monte Carlo results suggest that ignoring possible congestion and using standard logit errors can significantly bias estimates of price elasticities and welfare effects, even in random-coefficient, BLP-style models.

5. Empirical example

We end with an empirical example. Rysman (2004) studies a dataset on the Yellow Pages industry, measuring the positive feedback loop between consumers’ choice of directory (which is driven by the amount of advertising in the directory) and retailers’ placement of advertisements in directories (which is driven by consumer usage patterns). Rysman models the consumer’s decision as a discrete choice between available directories and an unspecified outside option. He observes a cross-section of directories and usage behavior where consumers in different geographic markets have access to different numbers of directories. Figure 1 shows the percentage of consumers served by different numbers of directories. The variance in this number of directories makes this a natural place to apply the techniques presented in this article.\textsuperscript{25} Correctly estimating the elasticity of usage to the quantity of advertising in a directory is important for measuring the importance of the feedback loop. In addition, correctly measuring the welfare benefits of competing directories is important for the policy question studied in the article.\textsuperscript{26}

The dataset consists of observations on the number of uses, per household, per month, in the distribution areas of 428 directories in 1996.\textsuperscript{27} We assume that a representative consumer needs information of the kind she could find in the Yellow Pages $M$ times per month. The exogenous parameter $M$ is constant across markets. Each time a consumer needs information, she can use one of the Yellow Pages in the area or turn to the outside option. The utility to consumer $i$ from using directory $j$ is

\[ u_{ij} = \beta_1 \ln(A_j) + X_j \beta_2 + \xi_j + \epsilon_{ij}. \]

The variable $A_j$ is the quantity of advertising at directory $j$, and the matrix $X_j$ represents demographic variables that may affect usage.\textsuperscript{28} The variable $\xi_j$ represents directory-specific factors that are unobservable to the econometrician, such as the quality of the book or regional usage habits.

We estimate this model and a model with our adjustment. A complicating factor is that Yellow Pages distribution areas overlap with each other. A directory may face no competitors for some of its consumers and one or more competitors for another group of consumers. Although we observe these distribution areas, we cannot distinguish how much usage comes from different portions of a directory’s distribution area.

Even so, implementing the simple logit model is straightforward. We observe $s_j$ (the market share for directory $j$) and $s_0$ (the market share for the outside option\textsuperscript{29}) in directory $j$’s total

\[ s_j + s_0 = 1. \]

\textsuperscript{24} That is, the effect that increasing the variance of $X$ will tend to makes the estimated model match the second comparative static better and the third comparative static worse.

\textsuperscript{25} Since Yellow Pages are not sold through retail stores, there is no literal retail congestion in this market. However, one can think of our congestion model as capturing the possibility that households have a limited amount of bookshelf or drawer space, and throw out books that don’t fit.

\textsuperscript{26} The policy question is whether or not welfare improves as competition increases. Multiple directories reduce market power but dissipate network effects. Rysman also estimates retailer demand for advertising and a publisher’s first-order condition for setting the quantity of advertising. Here, we focus only on the consumer’s decision.

\textsuperscript{27} The data were collected by National Yellow Pages Monitor. NYPM survey respondents maintain diaries of their Yellow Pages usage for one week. NYPM normally surveys between 1,000 and 3,000 people per MSA, although it used 11,200 respondents in the Los Angeles area. This usually results in at least a few hundred respondents even for very small directories.

\textsuperscript{28} As a measure of advertising, Rysman uses the number of pages in a book times the number of columns in a directory. The number is multiplied by $\delta$ for directories that are observably smaller than a standard directory. For $X_j$, each directory is associated with a central county, and $X_j$ comes from county-level census data.

\textsuperscript{29} We assume that $M = 26$. The highest number of uses per household in our dataset is 23.6, with an average of 11.4. The average for $s_0$ in our dataset is 47.7%.
market, and submarkets (areas of a directory’s market that are served by a uniform set of directories) are distinguished only by the presence of an “irrelevant alternative.” Under the logit model, the ratio $s_j/s_0$ is independent of the presence of these alternatives, so $s_j/s_0$ is the same in each submarket. Therefore, we can use the standard logit equation. For the simple logit model, we estimate

$$\ln(s_j) - \ln(s_0) = \alpha \ln(A_j) + X_j \beta + \xi_j.$$ 

To implement the crowding model, we take the crowding term to be the population weighted average of $R_j$ across submarkets. In that case, we estimate

$$\ln(s_j) - \ln(s_0) = \alpha \ln(A_j) + X_j \beta + \ln(R_j) + \xi_j,$$

where

$$R_j = \sum_{k \in K(j)} \psi_{jk} R_{J(k)}.$$ 

Here, $K(j)$ is the set of submarkets in $j$’s market area, $\psi_{jk}$ is the percentage of $j$’s population that lives in submarket $k$, and $J(k)$ is the number of products in submarket $k$.

We use two specifications of the crowding term $R_j$. The first is the parameterization suggested in Section 3: $R_J = -\left(\gamma + (1 - \gamma)J/J\right)$. The second specification is nonparametric; we allow the $R_J$ to take on different values for each $J$. We observe very few markets with more than 5 directories, so we restrict markets with 6, 7, or 8 directories to have the same adjustment parameter in the nonparametric case. We estimate both specifications by the generalized method of moments (Hansen, 1982) using the same set of instruments as in Rysman (2004).

Results appear in Table 3. Parameter estimates suggest that congestion is important. In the parametric case, $\gamma = .62$ and is precisely measured. Recall that $\gamma = 0$ implies no crowding and $\gamma = 1$ is full crowding. In the nonparametric case, the parameters for the crowding term are close to being monotonic in $J$ and decrease at a decreasing rate. Wald tests reject the joint equality of the estimates for different $J$. Regarding estimates of the other parameters, the two crowding models find coefficients closer to zero than the simple logit model, presumably to compensate for the effect of the crowding terms on elasticities.

Table 4 presents summary statistics. The columns on the left present elasticities of usage with respect to advertising. While differences across are not tremendously large, there are some differences between the models. First, it appears that the standard logit specification overestimates elasticities by 10% to 20%. Second, the standard logit model underpredicts changes in elasticities as the number of products increases. When the number of products goes from 1 to 8, the standard logit model shows that elasticity increases by 14%, whereas the crowding models both find that elasticity increases by 23%. This coincides with our intuition about how standard logit-based models restrict the extent to which crowding can occur as the number of products increases.

More striking are the welfare calculations. The logit model predicts that even the 7th and 8th Yellow Pages directories imply nontrivial welfare increases, over a third of what the first directory generates. On the other hand, the crowding model implies much lower benefits from new directories. When going from 1 to 8 directories, the standard model finds that welfare increases by over 400%. Under the crowding models, welfare increases by 180% and 146% for the parametric and nonparametric cases. These rates of increase are precisely measured and significantly different across models. Note that the nonparametric model actually finds that welfare decreases for when going from 3 to 4 directories. The possibility that welfare actually increases is well within confidence intervals for these estimates, and this result disappears when we parameterize the crowding function. For assessing welfare gains to new products, and to a lesser extent in estimating advertising elasticities, standard logit-based models appear to give biased results in this data.

© RAND 2005.
6. An alternative approach

In this section we briefly propose an alternative approach to allowing crowding in standard discrete-choice models. Intuitively, this approach tries to model a situation where additional firms entering the market differentiate into dimensions of unobserved characteristic space that consumers care less about. This seems to make intuitive sense; for example, in a market with only a few breakfast cereals, cereals may be primarily differentiated by how healthy they are or how crunchy they are. In a market with many cereals, cereals may be primarily differentiated only by the characters on their boxes, likely a less important characteristic. Ackerberg and Rysman (2003) (www.rje.org/main/sup-mat.html), construct a structural model exhibiting this property. Applied to a basic logit model, this model generates market shares of the form

\[ s_j = \frac{\exp \left( \frac{\beta_0 + \beta_1 X_j}{\mu(J, \tau)} \right)}{1 + \sum_{k=1}^{J} \exp \left( \frac{\beta_0 + \beta_1 X_k}{\mu(J, \tau)} \right)} \]

where \( \mu(J, \tau) \) is a function of the number of products in the market and a parameter \( \tau \). Note the similarity between this model and the model of Section 3. Both allow \( J \) to enter the market share equation: the former adjusts the equation multiplicatively, the latter adjusts it additively. This multiplicative adjustment essentially allows the variance of the logit errors to depend on \( J \). Ackerberg and Rysman (2003) show that the implications of this “multiplicative” approach are very similar to what we derived above for the “additive” approach. A \( \mu(J, \tau) \) that decreases in \( J \) implies that welfare benefits of new products in crowded markets are attenuated, that elasticities increase in more crowded markets (relative to a case without a crowding term), and that the three comparative statics from Section 2 can be matched with the additional parameter \( \tau \). They also discuss estimation and study the impact of the adjustment in Monte Carlo studies similar to those here.

7. Conclusion

This article highlights problems that arise as a result of the way that standard discrete-choice models handle symmetric unobserved product differentiation. We show that restrictive assumptions about the relationship between the number of products in a market and the dimensionality of unobserved characteristic space can lead to significantly biased estimates of elasticities and welfare changes. We suggest a straightforward adjustment that introduces the number of products in a market into the estimating equation. We present a structural interpretation of our solutions, showing how it could arise from the an agent maximization problem. We end with Monte Carlo and empirical evidence showing that this issue can be important in practice.

References


© RAND 2005.
### TABLE 1  Monte Carlo Results for Nested Logit Model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimator</th>
<th>Own-Price Elasticity</th>
<th>Cross-Price Elasticity</th>
<th>Outside-Good Price Elasticity</th>
<th>Welfare Increase (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau = 1$</td>
<td>Truth</td>
<td>$-2.29$</td>
<td>$0.21$</td>
<td>$0.11$</td>
<td>0</td>
</tr>
<tr>
<td>(Full Congestion)</td>
<td>Nested Logit</td>
<td>$-11.07$</td>
<td>$2.54$</td>
<td>$0.07$</td>
<td>20.00</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-2.28$</td>
<td>$0.22$</td>
<td>$0.12$</td>
<td>28.40</td>
</tr>
<tr>
<td>$\sigma = .2$</td>
<td>Nested Logit</td>
<td>$-4.85$</td>
<td>$0.97$</td>
<td>$0.09$</td>
<td>60.60</td>
</tr>
<tr>
<td>$\tau = .80$</td>
<td>Truth</td>
<td>$-2.25$</td>
<td>$0.25$</td>
<td>$0.15$</td>
<td>94.30</td>
</tr>
<tr>
<td>$\tau = .5$</td>
<td>Truth</td>
<td>$-2.98$</td>
<td>$0.49$</td>
<td>$0.13$</td>
<td>146.40</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-2.42$</td>
<td>$0.36$</td>
<td>$0.18$</td>
<td>233.20</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Nested Logit</td>
<td>$-8.28$</td>
<td>$1.71$</td>
<td>$0.11$</td>
<td>6.50</td>
</tr>
<tr>
<td>$\sigma = .2$</td>
<td>Nested Logit</td>
<td>$-1.39$</td>
<td>$2.29$</td>
<td>$0.10$</td>
<td>30.90</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-2.18$</td>
<td>$0.31$</td>
<td>$0.21$</td>
<td>23.00</td>
</tr>
<tr>
<td>$\beta_0 = 2$</td>
<td>Nested Logit</td>
<td>$-4.95$</td>
<td>$1.04$</td>
<td>$0.16$</td>
<td>48.20</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-4.75$</td>
<td>$0.24$</td>
<td>$0.04$</td>
<td>33.50</td>
</tr>
<tr>
<td>$\beta_1 = -2$</td>
<td>Nested Logit</td>
<td>$-7.87$</td>
<td>$1.26$</td>
<td>$0.03$</td>
<td>110.10</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-1.09$</td>
<td>$0.16$</td>
<td>$0.11$</td>
<td>23.00</td>
</tr>
<tr>
<td>Mean($X$) = 1</td>
<td>Nested Logit</td>
<td>$-2.39$</td>
<td>$0.52$</td>
<td>$0.08$</td>
<td>48.20</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-2.28$</td>
<td>$0.22$</td>
<td>$0.12$</td>
<td>28.40</td>
</tr>
<tr>
<td>Var($X$) = .95</td>
<td>Nested Logit</td>
<td>$-3.51$</td>
<td>$0.61$</td>
<td>$0.10$</td>
<td>110.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimator</th>
<th>Own-Price Elasticity</th>
<th>Cross-Price Elasticity</th>
<th>Outside-Good Price Elasticity</th>
<th>Welfare Increase (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-4.31$</td>
<td>$0.09$</td>
<td>$0.08$</td>
<td>5.1</td>
</tr>
<tr>
<td>$\tau = .80$</td>
<td>RCM</td>
<td>$-9.69$</td>
<td>$1.7$</td>
<td>$0.07$</td>
<td>29.20</td>
</tr>
<tr>
<td>$\tau = .5$</td>
<td>RCM</td>
<td>$-4.28$</td>
<td>$0.12$</td>
<td>$0.11$</td>
<td>19.40</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>RCM</td>
<td>$-5.66$</td>
<td>$0.62$</td>
<td>$0.11$</td>
<td>41.90</td>
</tr>
<tr>
<td>$\tau = .5$</td>
<td>RCM</td>
<td>$-4.23$</td>
<td>$0.19$</td>
<td>$0.17$</td>
<td>44.10</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>RCM</td>
<td>$-4.6$</td>
<td>$0.35$</td>
<td>$0.17$</td>
<td>58.40</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-4.55$</td>
<td>$0.08$</td>
<td>$0.06$</td>
<td>3.80</td>
</tr>
<tr>
<td>$\sigma = .4$</td>
<td>RCM</td>
<td>$-1.14$</td>
<td>$1.76$</td>
<td>$0.05$</td>
<td>22.60</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-4.06$</td>
<td>$0.38$</td>
<td>$0.35$</td>
<td>14.50</td>
</tr>
<tr>
<td>$\beta_0 = 4$</td>
<td>RCM</td>
<td>$-9.34$</td>
<td>$1.89$</td>
<td>$0.28$</td>
<td>62.30</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-6.38$</td>
<td>$0.02$</td>
<td>$0.02$</td>
<td>0.80</td>
</tr>
<tr>
<td>$\beta_1 = -3$</td>
<td>RCM</td>
<td>$-14.21$</td>
<td>$2.69$</td>
<td>$0.01$</td>
<td>6.10</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-2.05$</td>
<td>$0.21$</td>
<td>$0.21$</td>
<td>15.40</td>
</tr>
<tr>
<td>Mean($X$) = 1</td>
<td>RCM</td>
<td>$-3.65$</td>
<td>$0.69$</td>
<td>$0.17$</td>
<td>66.20</td>
</tr>
<tr>
<td>$\tau = .95$</td>
<td>Truth</td>
<td>$-4.31$</td>
<td>$0.09$</td>
<td>$0.08$</td>
<td>5.10</td>
</tr>
<tr>
<td>Var($X$) = .95</td>
<td>RCM</td>
<td>$-6.13$</td>
<td>$0.76$</td>
<td>$0.07$</td>
<td>30.80</td>
</tr>
</tbody>
</table>

RCM = Random-Coefficients Model.
### TABLE 3  
Estimation Results for Yellow Pages Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Standard Coefficient</th>
<th>Standard Error</th>
<th>Coefficient Crowding Term</th>
<th>Standard Error</th>
<th>Coefficient Crowding Term</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advertising</td>
<td>.705 (.069)</td>
<td></td>
<td>.631 (.070)</td>
<td></td>
<td>.632 (.073)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>−6.08 (1.07)</td>
<td></td>
<td>−4.92 (1.01)</td>
<td></td>
<td>−4.94 (1.17)</td>
<td></td>
</tr>
<tr>
<td>% urban population</td>
<td>−.023 (.006)</td>
<td></td>
<td>−.016 (.005)</td>
<td></td>
<td>−.013 (.005)</td>
<td></td>
</tr>
<tr>
<td>% lived in different county</td>
<td>.078 (.015)</td>
<td></td>
<td>.058 (.013)</td>
<td></td>
<td>.061 (.016)</td>
<td></td>
</tr>
<tr>
<td>% lived in different state</td>
<td>.047 (.020)</td>
<td></td>
<td>.031 (.017)</td>
<td></td>
<td>.027 (.023)</td>
<td></td>
</tr>
<tr>
<td>% own house</td>
<td>−.019 (.012)</td>
<td></td>
<td>−.020 (.011)</td>
<td></td>
<td>−.021 (.012)</td>
<td></td>
</tr>
<tr>
<td>% graduated high school</td>
<td>−.042 (.014)</td>
<td></td>
<td>−.032 (.012)</td>
<td></td>
<td>−.040 (.013)</td>
<td></td>
</tr>
<tr>
<td>% graduated college</td>
<td>−.015 (.016)</td>
<td></td>
<td>−.023 (.014)</td>
<td></td>
<td>−.007 (.016)</td>
<td></td>
</tr>
<tr>
<td>Per-capita income</td>
<td>.029 (.021)</td>
<td></td>
<td>.032 (.018)</td>
<td></td>
<td>.022 (.022)</td>
<td></td>
</tr>
<tr>
<td>Telco book</td>
<td>1.156 (.103)</td>
<td></td>
<td>1.050 (.100)</td>
<td></td>
<td>1.018 (.103)</td>
<td></td>
</tr>
<tr>
<td>County population growth</td>
<td>.003 (.016)</td>
<td></td>
<td>.012 (.014)</td>
<td></td>
<td>.016 (.015)</td>
<td></td>
</tr>
<tr>
<td>% take public transporation</td>
<td>−.035 (.030)</td>
<td></td>
<td>−.023 (.027)</td>
<td></td>
<td>−.041 (.034)</td>
<td></td>
</tr>
<tr>
<td>% have not moved</td>
<td>.072 (.016)</td>
<td></td>
<td>.047 (.015)</td>
<td></td>
<td>.054 (.019)</td>
<td></td>
</tr>
<tr>
<td>Population density</td>
<td>−1.11E-0 (3.88E-05)</td>
<td></td>
<td>−8.39E-05 (3.43E-05)</td>
<td></td>
<td>−7.20E-05 (3.50E-05)</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>.616 (.120)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adjustment

- $J = 1$ : 0  
- $J = 2$ : −.350 (.142)  
- $J = 3$ : −.343 (.177)  
- $J = 4$ : −.743 (.217)  
- $J = 5$ : −.865 (.308)  
- $J = 6, 7, 8$ : −.967 (.364)
TABLE 4  Summary Variables for Yellow Pages Data

<table>
<thead>
<tr>
<th>Firms</th>
<th>Elasticity</th>
<th></th>
<th></th>
<th></th>
<th>Firms’ Welfare</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard</td>
<td>Parametric</td>
<td>Nonparametric</td>
<td></td>
<td>Standard</td>
<td>Parametric</td>
<td>Nonparametric</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.55 (.052)</td>
<td>.45 (.053)</td>
<td>.45 (.054)</td>
<td>.20 (.007)</td>
<td>.28 (.025)</td>
<td>.27 (.026)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.58 (.056)</td>
<td>.52 (.057)</td>
<td>.52 (.060)</td>
<td>.36 (.012)</td>
<td>.37 (.012)</td>
<td>.36 (.026)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.60 (.058)</td>
<td>.55 (.059)</td>
<td>.54 (.060)</td>
<td>.51 (.015)</td>
<td>.45 (.019)</td>
<td>.50 (.044)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.61 (.059)</td>
<td>.56 (.060)</td>
<td>.57 (.064)</td>
<td>.63 (.018)</td>
<td>.52 (.032)</td>
<td>.46 (.061)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.62 (.060)</td>
<td>.57 (.061)</td>
<td>.58 (.066)</td>
<td>.74 (.020)</td>
<td>.59 (.044)</td>
<td>.50 (.108)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.63 (.061)</td>
<td>.57 (.062)</td>
<td>.58 (.066)</td>
<td>.84 (.022)</td>
<td>.66 (.055)</td>
<td>.53 (.136)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.64 (.062)</td>
<td>.58 (.063)</td>
<td>.59 (.066)</td>
<td>.93 (.023)</td>
<td>.72 (.064)</td>
<td>.60 (.149)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.64 (.062)</td>
<td>.58 (.063)</td>
<td>.59 (.067)</td>
<td>1.02 (.024)</td>
<td>.78 (.073)</td>
<td>.66 (.159)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase (%)</td>
<td>410.5 (5.4)</td>
<td>180.5 (49.4)</td>
<td>146.1 (67.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© RAND 2005.