NBER WORKING PAPER SERIES

UNOBSERVED PRODUCT DIFFERENTIATION IN DISCRETE CHOICE MODELS: ESTIMATING PRICE ELASTICITIES AND WELFARE EFFECTS

Daniel A. Ackerberg
Marc Rysman

Working Paper 8798
http://www.nber.org/papers/w8798

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2002

Unobserved Product Differentiation in Discrete Choice Models:
Estimating Price Elasticities and Welfare Effects
Daniel A. Ackerberg and Marc Rysman
NBER Working Paper No. 8798
February 2002
JEL No. L10, C25

ABSTRACT

Standard discrete choice models such as logit, nested logit, and random coefficients models place very strong restrictions on how unobservable product space increases with the number of products. We argue (and show with Monte Carlo experiments) that these restrictions can lead to biased conclusions regarding price elasticities and welfare consequences from additional products. In addition, these restrictions can identify parameters which are not intuitively identified given the data at hand. We suggest two alternative models that relax these restrictions, both motivated by structural interpretations. Monte-Carlo experiments and an application to data show that these alternative models perform well in practice.

Daniel A. Ackerberg
Department of Economics
UCLA
8283 Bunche Hall
Los Angeles, CA 90095
and NBER
ackerber@econ.ucla.edu

Marc Rysman
Boston University
Boston, MA 02215

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 differentiated characteristics that are observed by the econometrician. In addition, these models also typically allow for unobserved product characteristics through the inclusion some form of "symmetric unobserved product differentiation" (SUPD)[1]

The most common example of SUPD are logit errors in consumers' utility functions (see McFadden, 1974). These logit errors represent unobserved (to the econometrician) product differentiation that is symmetric across products. 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 differentiation often prevents these models from predicting zero market shares, an obviously desirable feature. Its inclusion can also ease estimation.

This paper argues that while SUPD in itself may be helpful, standard models (e.g. logit models, probit models, nested logit models, and random coefficient models) implement it in an undesirable way. These models assume that the number of products or alternatives available in a market and the dimensionality of SUPD are linked in an extremely particular way. Specifically, each product added to the market adds one additional dimension to SUPD space. This results in very little "congestion" in unobserved characteristic space and can be problematic in situations where different consumers face different numbers of products, either because consumers are drawn from different geographies or from different time periods.[2] Researchers may intuitively think that in markets with more products, characteristic space should "fill up" in some sense. These standard models place strong restrictions on how this occurs with regards to unobservable

[1]Notable exceptions are Bresnahan (1987) and Feenstra and Levinsohn (1994).

[2]There are many examples in the literature. Berry and Waldfogel (1999) and Rysman (2002) face cross-sectional variation in the number of available products. Berry, Levinsohn and Pakes (1995), Bresnahan, Stern and Trajtenberg (1997) and Petrin (1999) face temporal variation. Nevo (2002) and Shum (1998) face both. Arcidiacono (2000) studies high school students choosing colleges after acceptance letters have been received, so his "consumers" face different number of "products" because of an institutional process. This list is very far from exhaustive.

characteristics.

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 (1999) and Berry and Pakes (1999). Because of the lack of crowding in the standard treatment of SUPD, welfare calculations in these 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, both in terms of magnitudes and statistical significance. The basic idea here is that the restrictions of standard SUPD force variation in the number of products in the choice set to identify (or help identify) price elasticites. Interestingly, we show that with these restrictions, one can "identify" price elasticites without ever observing any variation in prices. We feel that this source of identification is ad-hoc since it relies completely on the precise assumption that there is very little congestion in unobserved product characteristic space. This source of identification is even more unreliable if, as is often the case, "defining" different products has some arbitrariness to it.[3]

There are two previous approaches in the literature that address these issues. The first set of work (e.g. Berry, Levinsohn and Pakes (1995) - henceforth BLP, and Petrin) 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). The basic idea is to keep SUPD (e.g.. logit errors) in the model but, by allowing significant amounts of heterogeneity in tastes around *observed* product characteristics, attempt to reduce SUPD's importance. These methodologies have worked to varied extents in reducing the influence of SUPD – success is proportional to the econometrician's ability to *observe* the relevant differentiated characteristics. However, as inflexible SUPD still exists in these models, its effects still exist.

A second and more recent approach, advocated by Berry and Pakes (1999) and Bajari and Benkard (2001), eliminates SUPD altogether from the model. In their "pure hedonic" models, products are un-

[3]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 vs (BMW 330, BMW 325) vs. (BMW 330i, BMW 330Ci, BMW 330 Ci Convertible).

observably differentiated only with respect to a single dimensional unobserved characteristic.[4] As new products enter, this unobserved product space fills up. We envision a couple of potential limitations of this type of model. First, this approach might be unreasonably restrictive in the opposite direction from the standard models. While there is a sense that unobserved product space expands too much with logit errors, there is a sense that it expands too little with the pure hedonic models. It may be restrictive to disallow new products from expanding unobserved characteristic space (e.g. differentiate in new dimensions). Again we expect this restriction to be less significant as the econometrician is able to observe more of the relevant characteristics - this will depend on the empirical application. These models can also be more complex to estimate than models including a logit error.

This paper suggests a third approach, which we interpret as somewhat of a compromise between the above two. We argue that it is only the *unnecessary inflexibility* of standard SUPD that can adversely affect parameters of interest such as substitution patterns and welfare effects. As such, we keep SUPD in our model, but allow the SUPD to be considerably more flexible than is currently used. This flexibility allows an econometrician to *estimate* how fast unobserved product space expands with the addition of new products, not *assume it*, as prior work does.

In essence, our approach allows functions of the number of products in a market (and/or the number of products in a group or nest) to enter the discrete choice estimating equation. While this might initially seem ad hoc, we show that each of these models have an intuitive and realistic structural interpretation.

The first structural model is one of retail product congestion. Products in this model are sold through a retail market with a limited number of stores. As new products enter the market, they can "crowd out" existing products from retail stores. This model generates an additive adjustment to the estimating equation which is a function of the number of products. The second model allows the variance of the logit error to be smaller in markets with more products. We show how this feature can arise from a model in which products in crowded markets differentiate into dimensions that consumers care less about. This model generates a multiplicative adjustment of the estimating equation, also a function of the number of products.

We proceed as follows. In Section 2 we argue that 1) traditional discrete choice models place unnecessary restrictions on SUPD, 2) that these restrictions can "identify" parameters that intuitively should not

[4] Feenstra and Levinsohn (1995) also estimate a multidimensional pure hedonic model, albeit without any unobserved characteristics.

be identified, and 3) that these restrictions can bias parameters of interest. Section 3 introduces our two models of product congestion and discusses their estimation. In Section 4, we present Monte-Carlo results which show 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 our techniques to a data set previously used in Rysman (2002).

Lastly, note that much of our applications are focused on the context of estimating aggregated discrete choice models. The reason that we focus on aggregated discrete choice models is that these tend to be estimated on data across markets (in space or time) where one often observes changes in the size of the choice set. As our technique is expressly for dealing with such changes, this is where it is most applicable and probably most 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 first argues that error structures used in traditional discrete choice models are unnecessarily restrictive, which leads to undesirable identification results. Second, the section shows that these restrictions have adverse affects on parameters of interest such as price elasticities and welfare calculations. We also briefly suggest our solutions to the problem, though this is formalized and further motivated in Section 3. Throughout, we use the nested logit model to illustrate our points. However, our arguments extend to other discrete choice models as well.[5]

2.1 Identification

We use 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 nest g and the outside option is in a nest by itself. In the nested

[5]In particular, it applies to random coefficients models. Note that the nested logit model is a special case of a random coefficients model where random coefficients are only on group dummy variables.

logit model, the utility obtained by consumer i from product j ($j > 0$) is:

$$u_{ij} = \beta_0 + X_j\beta_1 + \zeta_{ig} + \epsilon_{ij}$$

where ϵ_{ij} is distributed Extreme Value with variance $(\pi\mu_2)^2/3$, ζ_{ig} is constant for each individual across the product nest and ζ_{ig} is distributed such that $\zeta_{ig} + \varepsilon_{ij}$ is distributed Extreme Value with variance $(\pi\mu_1)^2/3$. As is standard, we assume $u_{i0} = \zeta_{i0} + \epsilon_{i0}$, normalizing the mean utility of the outside option to 0. The variance scale parameters μ_1 and μ_2 are not separately identified but the ratio μ_2/μ_1 is, so it is useful to define the parameter $\sigma = \mu_2/\mu_1$, and normalize $\mu_2 = 1$. 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.[6]

This model implies that the within-group market share for j is:

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

Letting $D = \sum_{k=1}^{J}\exp(\beta_0 + X_k\beta_1)$, the group and total market shares are:

$$s_g = \frac{D^\sigma}{1 + D^\sigma} \qquad s_j = s_{j|g}\, s_g$$

Researchers observe 3 forms of variation under the nested logit model. The first type is variation in within-group market shares due to changes in observable product characteristics. Looking at this derivative tells us what parameters are identified by this type of variation. That comparative static is:

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

Therefore, this type of variation identifies β_1. The second type of variation is variation in group market shares due to changes in observable product characteristics. The third type of variation is variation in group shares due to changes in the number of products. In order 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:

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

[6]We ignore endogeneity issues regarding. price, which has been a focus of the prior literature. These issues are completely independent of the point we are making, which is valid whether price movements are purely exogenous or whether they are endogenous and one must find some exogenous source of price variation. We also follow the existing literature by assuming that product characteristics (other than price) and the number of products in the market is exogenous.

Therefore, there are two sources of identification for σ: cross-group switching from changes in the number of products and cross-group switching from changes in observed characteristics. There are also two sources of identification for β_1: within-group switching from changes in observed characteristics and cross-group switching from changes in observed characteristics. Three comparative statics ($\frac{\partial s_g}{\partial X}$, $\frac{\partial s_g}{\partial J}$, and $\frac{\partial s_j}{\partial X_j}$) are captured by only two parameters (β_1 and σ), so the model implies a restrictive relationship between the effects.[7] The easiest way to see this restriction is to note that the nested logit model assumes that the ratio between $\frac{\partial s_g}{\partial X}$ and $\frac{\partial s_g}{\partial J}$ is $\beta_1 J$, but β_1 could be identified by $\frac{\partial s_{j|g}}{\partial X_j}$.

These features have perverse 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 σ and β_1. Therefore, a researcher can identify the effect of adding a product to the choice set *without ever* observing variation in the number of products. Even more unintuitively, one can identify cross-price elasticities of products within the group *without ever* observing changes in relative prices of the products (for an example of this, see the start of Section 3.1). In our Monte Carlo results, we show that even when there are "good" sources of identification, e.g. relative price variation to identify price elasticities, potentially spurious identification from changes in the number of products in the choice set can bias these elasticities. Lastly, note that one way to summarize the basic intuition here is that all of the parameters in standard discrete choice models can be identified by estimating only in markets 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.

2.2 Implications for Estimating Elasticities and Welfare

Why do standard discrete choice models identify effects that intuitively should not be identified? Because they make very restrictive assumptions about the relationship between unobservable characteristic space and the number of products. Specifically, standard discrete choice models assume that markets with a high number of products are no more crowded (in unobserved characteristic space) than markets with a small number of products. For instance, we can write utility in the nested logit model in terms of dummy variables for products (d_j):

[7] Note that the constant term β_0 is identified by the level of market shares relative to the outside good.

$$u_{ij} = \beta_0 + X_j\beta_1 + \zeta_{ig} + d_1\epsilon_{i1} + ... + \ d_J\epsilon_{iJ} \quad j > 0$$

One might expect a new product to crowd out the initial unobserved product space. But the Jth product differentiates in an entirely new dimension (that of d_J) which is associated with an entire new set of logit errors, so the dimensionality of unobserved product space expands with the addition of the new product.

An implication of this restriction is that all products are "equi-distant" from each other in unobserved characteristic space and this distance remains constant as the number of products in the market changes. Precisely, if one randomly chooses two products in each of two markets, the expected difference between u_{i1} and u_{i2} is the same regardless of the number of products in the markets (for ease, suppose that $X_j = X$ $\forall j$). This is counterintuitive in the following way. With classical product differentiation models such as the Hotelling model or the Salop circular model in mind, one would naturally expect products in markets with more products to be "closer" in characteristic space.[8] This restriction of logit based models ends up playing a strong role in identifying price elasticities, as exhibited in the previous section.

There are a couple of additional perverse implications resulting from the lack of crowding. First, we expect these models to relatively under-predict elasticities in markets with more products (as they assume away congestion in large markets). We examine this issue in Monte-Carlo experiments. There is also a problem valuing new products. Because there is no crowding, we expect valuations of new products to be overestimated. This point regarding welfare has been made in previous work (Petrin (1999), Trajtenburg (1990)) and is also exemplified in our Monte-Carlo experiments.[9]

2.3 Proposed Solutions

We now briefly suggest two adjustments to these logit based models. These adjustments allow the models to deal with product crowding in a much more flexible way, alleviating the overidentification discussed above.

[8] For example, 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 3 products in the market, the expected difference is 1/3*1 + 2/3*1/2 = 2/3, with 4 products it is 3/6*1/3 + 2/6*2/3 + 1/6*1 = 5/9, with 5 products it is 4/10*1/4 + 3/10*2/4 + 2/10*3/4 + 1/10*1 = 1/2.

[9] 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 the logit. Extensions of the additive and multiplicative adjustment to the CES model are available from the authors.

At the same time, our approach allows for the *estimation* of the rate at which the dimensionality of product space increases. For sake of clarity, we present both models in terms of the nested logit model, but either adjustment is applicable to other models, such as the logit and the random coefficients model.

In the *additive model*, we add a function $f(J;\gamma)$ with parameter γ to the term $\beta_0 + X_j\beta_1$. We show in the next section how an additive model with $f(J;\gamma)$ declining in J can arise from a model of retail crowding. In the additive case, the within-group share function is:

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

Now, the three comparative statics discussed above are:

$$\frac{\partial s_{j|g}}{\partial X_j} = \beta_1 s_{j|g}(1-s_{j|g}) \quad \frac{\partial s_g}{\partial X} = \sigma\beta_1 s_g(1-s_g) \qquad \frac{\partial s_g}{\partial J} = \sigma s_g(1-s_g)\left(\frac{1}{J} + f'(J;\gamma)\right)$$

The first 2 comparative statics are the same as before, but the third now depends on parameters in the new function. This feature gives the nested logit model the ability to match all of the observed variation. Now the parameter σ can be clearly interpreted as capturing cross-group variation due to variation in characteristics (such as price) while the parameters in the new function capture cross-group variation due to changes in the number of products.

In the *multiplicative model*, we allow the variance of the unobservable portion of utility to depend on the number of products. In the nested logit model, this means defining $\mu_2 = \mu_2(J;\tau)$.[10] If $\mu_2'(J;\tau) < 0$, products in crowded markets are in a sense closer together. Equivalently, additional products are differentiated into dimensions that consumers care less about. We formalize this point in the next section. In the multiplicative model, the within-group market share function is:

$$s_{j|g} = \frac{\exp\left(\frac{\beta_0 + X_j\beta_1}{\mu_2(J;\tau)}\right)}{\sum_{k=1}^{J}\exp\left(\frac{\beta_0 + X_k\beta_1}{\mu_2(J;\tau)}\right)}$$

As with the additive model, parameters in $\mu_2(J;\tau)$ give the model the extra lever required to match the three comparative statics.

Now consider the effects of the additive and multiplicative adjustments on welfare and elasticity calculations. Clearly, estimating $f'(J;\gamma) < 0$ would allow the additive model to find smaller welfare benefits

[10] If μ_2 depends on J, then μ_1 (and σ) does also. We address this issue in Section 3.4

as the number of products increases. Similarly, the multiplicative adjustment allows for attenuated welfare benefits from high numbers of products. Also, these adjustments affect elasticities. In particular, allowing crowding (either $f'(J;\gamma) < 0$ or $\mu_2'(J;\tau) < 0$) results in greater increases in elasticities as markets become crowded then those implied by standard models.

3 A Structural Interpretation

In this section, we exhibit structural models that generate the adjustments suggested in the previous section. Doing so 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). First, we show how the additive adjustment can arise from a model of retail congestion. Second, we show how the multiplicative adjustment can arise from a model in which products in crowded markets differentiate into dimensions that consumers care less about.

3.1 A Model of Product Congestion

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 outlets 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 strong 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) is 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 product space (in this case franchise 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 with such congestion well - 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 (or the multiplicative model introduced below) to the fast food data described above - price elasticities would **not** be identified - an intuitive outcome given the lack of any variation in prices.

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 is **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 following form:

$$U_{ijr} = u_j + \epsilon_{ir}$$

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, ξ_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 ϵ_{ir}, not ϵ_{ij}. Intuitively, ϵ_{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 ϵ_{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 product 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:

$$s_j = \frac{R_j e^{u_j}}{1 + \sum_k R_k e^{u_k}} \tag{1}$$

$$= \frac{e^{u_j + \ln(R_j)}}{1 + \sum_k e^{u_k + \ln(R_k)}} \tag{2}$$

3.2 Estimating the Additive Model

For individual level data, 1 could be estimated by maximum likelihood. For aggregate data, this model can be estimated using the Berry (1994) inversion:

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

In practice, one needs to parametrically specify R_j. In the simplest case, where each product is sold in an equal number of stores, we have $R_j = \frac{R}{J}$ and we only need to 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 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, 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). Also, this specification could be estimated in the aggregate case by linear techniques. A drawback is that this specification lacks a clear structural interpretation of the parameter.

Lastly, one might estimate $R(J)$ non-parametrically. Given that J is discrete, this is extremely simple - one just includes indicator functions for different market size (with a normalization for one J).

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

$$R_j = f(J)e^{X_j\tau_1 - \tau_2 p_j + \tau_3 \xi_j}$$

so that product characteristics do affect R_j. In this case, for example, τ_1 is not separately identified from β. With other specifications of R_j, the different effects might be identified computationally, but this identification would be completely dependent on non-linearities. 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.

$$U_{ijr} = u_j + \epsilon^1_{ij} + \rho\epsilon^2_{ijr}$$

ϵ^1_{ij} is consumer i's product specific taste, ϵ^2_{ijr} is consumer i's product-retail outlet specific taste, and ρ 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 (

ϵ_{ijr}^2 distributed Type I Extreme Value, ϵ_{ij}^1 distributed such that $\epsilon_{ij}^1 + \rho\epsilon_{ijr}^2$ distributed Type I Extreme Value), we get the following product level market shares:

$$s_j = \frac{\left[R_j \exp(\frac{u_j}{\rho})\right]^\rho}{1 + \sum_k \left[R_k \exp(\frac{u_k}{\rho})\right]^\rho} = \frac{\exp(u_j + \rho \ln(R_j))}{1 + \sum_k \exp(u_k + \rho \ln(R_k))}$$

where R_j is the number of retail stores that product j is sold at. Again assuming all product are sold at an equal number of stores and that

$$R = \gamma + (1 - \gamma)J$$

we get

$$s_j = \frac{\exp(u_j + \rho \ln(\gamma/J + 1 - \gamma))}{1 + \sum_k \exp(u_k + \rho \ln(\gamma/J + 1 - \gamma))}$$

which leads to the estimating equation:

$$\ln(\frac{s_j}{s_0}) = u_j + \rho \ln(\gamma/J + 1 - \gamma)$$

Note that ρ and γ are formally separately identified in this model, but this separate identification is due to non-linearities in the J term and might be unreliable in practice. For instance, consider the specification $\ln(R_j) = \gamma \ln(J)$. Then the estimating equation is:

$$\ln(\frac{s_j}{s_0}) = u_j + \rho\gamma \ln(J)$$

where only the product $\rho\gamma$ is identified. Clearly, with a non-parametric specification of $R(J)$, ρ is also unidentified. Note that this lack of identification is not a bad thing. It simply means that our model is robust to unobserved tastes at both the product and retail store level. Separating the parameters (e.g. ρ vs. γ) is irrelevant for empirical or welfare implications.

This congestion model is easily generalizable 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 ζ_{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). Product shares in this model are given by:

$$s_j = s_{j|g} s_g = \frac{R_j e^{\frac{u_j}{\sigma}}}{\sum_{k \in g_j} R_k e^{\frac{u_k}{\sigma}}} \frac{\left(\sum_{k \in g_j} R_k e^{\frac{u_k}{\sigma}}\right)^{\sigma}}{1 + \sum_g \left(\sum_{k \in g} R_k e^{\frac{u_k}{\sigma}}\right)^{\sigma}}$$

and estimation can proceed using the Berry inversion:

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

One issue in the nested logit model is how to specify R_j. The number of retail outlets per product could be a function of the number of products in the nest, the total number of products in the market or some weighted average of the two. Our model is similarly adaptable to multiple level nested logit models, other GEV models (e.g. the model of Bresnahan, Stern and Trajtenberg) and random coefficients models.[11]

3.3 Variance in Discrete Choice Models

This subsection presents a structural justification of the multiplicative model. For motivation, consider the evolution of the market for ready-to-eat breakfast cereals. Initally, the market contained only a few products, and differentiation was across fundamental and likely very important features such as healthiness and taste. Recently, with so many many new products, it is likely that some products are distinguishable only by the colors on their box. The basic idea here is that as more products enter the market, they differentiate into dimensions (e.g. color of box) that are less important to consumers. This section shows that if we allow for this type of effect in unobserved characteristic space, we end up with our multiplicative model.

Standard discrete choice models imply that each product differentiates into a separate dimension, and that each dimension is equally important. Our innovation is to adjust the model so that products in crowded markets differentiate into dimensions that matter less to consumers. As a result, consumers are more responsive to changes in observable (to the econometrician) characteristics such as price in a crowded market, and the welfare from the last product is much lower in a crowded market.

[11] For random coefficients models, one could either 1) simply include the total number of products in the estimating equation (in essence assuming congestion occurs equally across products), or 2) extend the intution from the nested logit model described above. Instead of counting the number of products in the same nest, one could count the number of products weighted by how close they are in characteristic space.

An impediment to developing this model is that important concepts for analyzing product differentiation, such as the distance between products and travel costs for consumers, are not explicit in models such as the logit and probit. In contrast, these concepts are explicitly specified in an address (Hotelling) model. Therefore, our strategy is to specify a generalized empirical model and then an address model, and then present conditions such that the two models have the same implications for market shares. We then impose the features we want on the address model and show how those features lead to a tractable adjustment in the empirical model.

Anderson, De Palma and Thisse (1992), ADT, present an algorithm for linking an address model to a logit model.[12] By link, we mean that the models match each other in terms of market shares and elasticities to the mean utilities. Here, we extend their model for our purposes. We define the logit model as follows: A unit mass of agents choose 1 of $J+1$ products (which can be thought of as J products and an outside option). Each product is defined by quality level u_j. Each agent i receives utility level u_{ij} from a given product defined by $u_{ij} = u_j + \epsilon_{ij}$, where $\epsilon_{i0} \ldots \epsilon_{iJ}$ is a random variable drawn from an extreme value distribution with variance scale parameter μ. Each agent chooses the product that confers the highest utility, so the market share for product j is:

$$s_j = \frac{\exp(u_j/\mu)}{\Sigma_{k=0}^{J} \exp[u_k/\mu]}$$

Now we turn to specifying the *address model* corresponding to this logit model. There are $J+1$ distinct products, each characterized by a vertical utility u_j and a vector of characteristics $z_j \in \mathbb{R}^L$ over which consumers have idiosyncratic tastes. Each consumer i is characterized by a vector $c_i \in \mathbb{R}^L$ that describes the consumer's ideal product. Let the function $\tau(l)$ represent the consumer cost of travel in dimension l. We assume $l' > l'' \Rightarrow \tau(l') \leq \tau(l'')$, so location in higher dimension is less important. A consumer located at c_i who consumes product j receives utility level:

$$\widehat{u}_{ij}(c_i) = u_j - \sum_{l=1}^{L} \tau(l)(c_i^l - z_j^l)^2 \quad j = 0, \ldots, J.$$

ADT assume that travel costs are constant across dimensions and previous empirical work does so as well

[12]In fact, ADT present a general algorithm for linking an address model to any linear random utility model of discrete choice. Our adjustments are extendable to more other models, but all inuition is clear from the logit case.

(at least implicitly). Allowing travel cost to depend on the dimension is the structural change that we use to generate a more flexible discrete choice model suitable for estimation.

Consumers are distributed in $\mathbb{R}^L$ according to the probability density $g(c_i)$. Consumers choose the option that confers the most utility. Therefore, the market share of product j is:

$$\begin{aligned}
\widehat{s}_j &= \int_{\widehat{M}_j} g(c_i)dc_i, \quad j = 0, \ldots, J. \\
\text{where } \widehat{M}_j &= \{c_i \in \mathbb{R}^L | \widehat{u}_{ij}(c_i) = \max_{k=0,\ldots,J}(\widehat{u}_{ik}(c_i))\}
\end{aligned}$$

Note that $\sum_{j=0}^{J} \widehat{s}_j = 1$. We seek assumptions such that:

Condition 1 (Match) $s_j = \widehat{s}_j$ *and* $\frac{\partial s_j}{\partial u_k} = \frac{\partial \widehat{s}_j}{\partial u_k}$ $j, k = 0, \ldots, J$

Satisfying Condition 1 requires specifying how the extreme value distribution pins down consumer and product locations in the address model. As Section 2.2 points out, the idiosyncratic portion of utility in the logit model can be thought of as a vector of product-specific dummy variables interacted with the consumer's vector ϵ_i. In the address model, we use the vector of dummy variables to create product locations, and the vector ϵ_i to create consumer locations. To begin, we assume that the number of product characteristics in the address model is equal to the number of products, i.e. $L = J$. Then, product locations are specified as follows:

Assumption 1
$$\begin{aligned}
z_j^l &= \begin{cases} b & \text{if } l = j, \quad j, l = 1 \ldots J \\ -b & \text{otherwise} \end{cases} \\
z_0^l &= -b \quad l = 1, \ldots, J.
\end{aligned}$$

Products are located at positions such as $\{-b, -b, \ldots, b, \ldots, -b, -b\} \in \mathbb{R}^J$. The parameter b measures the proximity of products. The specification mimics the vector d_j but with the advantage (over something like $\{0, 0, \ldots, b, \ldots, 0, 0\}$) that consumers who are indifferent between products are located on the axis. This simplifies notation in specifying consumer locations.

Given product locations and the consumer utility function, specifying the distribution of consumers defines the address model. First, consider the case of $\tau(l) = \tau \ \forall l$. ADT show that for this case, Condition 1

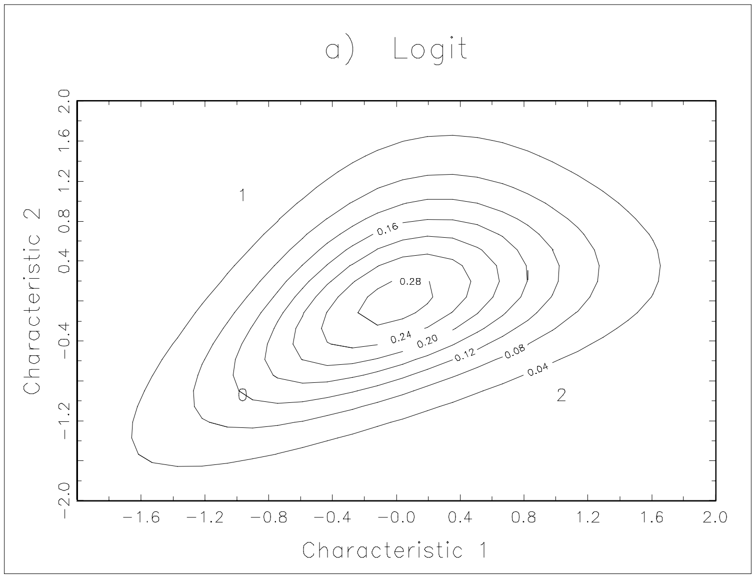

Figure 1: Consumer Distribution in the Address Model that Matches a Logit Model

is satisfied if:

$$g(c_i) = \left(\frac{4b\tau}{\mu}\right)^J (J)! \frac{\Pi_{j=1}^{J} \exp[4b\tau(c_i^j - c_i^0)/\mu]}{\left(1 + \Sigma_{j=1}^{J} \exp[4b\tau(c_i^j - c_i^0)/\mu]\right)^J} \tag{3}$$

where $c(u) : \mathbb{R}^{J+1} \to \mathbb{R}^J$ is such that $c^j(u) = (u^j - u^0)/4b\tau(j)$

A few features of the model bear comment. Travel costs (τ) and product distances (b) enter in the same way. Not surprisingly, a given distribution of consumers could generate the same market shares either because products are distant from each other or because travel costs are high. Also, μ has the inverse role of τ and b. That is, for a given set of market shares and elasticities, high variance of ε in the logit model is accounted for by high travel costs or distant products in the address model. Finally, the lack of crowding is made explicit in Assumption 1. Each product is equidistant from the outside option and equidistant from each other, regardless of how many products there are.

For further insight into the model, consider the $J = 3$, $m = 2$ case. Figure 1 draws a contour map of $g(c)$ for $b = 1$, $\tau = 1$ and $\mu = 2$. Contour lines form an approximation of an equilateral triangle in between each product. The graph makes it clear how little is pinned down by linking the address model to the empirical model. For instance, for a different set of parameters b, τ, and μ, we simply compute a different distribution of consumers and the implications for market shares are unchanged. This gives some leeway in modelling how the environment changes as J increases.

Now consider our adjustment, that $\tau(l)$ decreases in l. Given Assumption 1, each product is differentiated into a distinct dimension so each product j can be associated with a separate travel cost $\tau(j)$. The assumption that $\tau(j)$ decreases in j means that products with high j are differentiated into a dimension that consumers do not value highly. These products add very little to total welfare and have very high elasticities with respect to observable features (u_j) – exactly what we might expect in crowded markets.

The next question is, how can decreasing travel costs be represented in the logit model? In Equation 3, we would like to replace τ with $\tau(j)$ but have b and $g(\cdot)$ remain the same. From inspection, it is clear that Condition 1 can be satisfied if we allow μ to also depend on j. So the fact that some product's unobservable differentiation is in less important dimensions is captured in the logit model by having those products have lower variance in their unobservable utility. We replace τ with $\tau(j)$ and μ with $\widetilde{\mu}(j)$ and rewrite Equation 3 as:

$$g(c_i) = (4b)^J \prod_{j=1}^{J} \left(\frac{\tau(j)}{\widetilde{\mu}(j)}\right)(J)! \frac{\Pi_{j=1}^{J} \exp[4b\tau(j)(c_i^j - c_i^0)/\widetilde{\mu}(j)]}{\left(1 + \Sigma_{j=1}^{J} \exp[4b\tau(j)(c_i^j - c_i^0)/\widetilde{\mu}(j)]\right)^{J+1}}$$

For the appropriately chosen $\widetilde{\mu}(j)$, the distribution $g(c_i)$ is unchanged. Using this equation as the link between the address model and the empirical model implies that the new logit share function is:

$$s_j = \frac{\exp(u_j/\widetilde{\mu}(j))}{1 + \sum_{k=1}^{J} \exp(u_k/\widetilde{\mu}(k))}$$

where $u_0 = 0$.

A major concern for estimating this share function is that it requires researchers to assign products to specific dimensions. Researchers are unlikely to want to make assumptions about something so abstract. A solution is to integrate over all possibilities (with equal weights). There are $J!$ possible sequences of J products in dimension space. Define $I : [1, J!] \times [1, J] \to [1, J]$ such that $I(m, j)$ give the location of choice j in sequence m. Then s_j can be written as:

$$s_j = \sum_{m=1}^{J!} \frac{\exp[u_j/\widetilde{\mu}(I(m, j))]}{1 + \sum_{k=1}^{J} \exp[u_k/\widetilde{\mu}(I(m, k))]} \frac{1}{J!}$$

This share function looks computationally burdensome. A further simplification is available by noting that this share function treats each product symmetrically. So there exists a function $\mu(J)$ such that:

$$s_j = \frac{\exp(u_j/\mu(J))}{1 + \sum_{k=1}^{J} \exp(u_k/\mu(J))} \tag{4}$$

3.4 Estimating the Multiplicative Model

As with the additive model, the multiplicative model can be estimated by maximum likelihood (typically for individual level data) or the Berry (1994) inversion:

$$\ln(s_j) - \ln(s_0) = \frac{u_j}{\mu(J)}$$

where J is the total number of products in j's market. Note that one needs to normalize $\mu(J)$ for some value of J and then parameterize $\mu(\cdot)$. One caveat is that a non-linear estimation technique is required to estimate this equation, but it is otherwise straightforward.

Interesting issues arise if the researcher would like to use this approach in a nested logit framework. Consider the model in Section 2.1. Writing out the market share accounting for μ_1 and μ_2 results in:

$$s_j = \frac{e^{u_j/\mu_2}}{\sum_{k=1}^{J} e^{u_k/\mu_2}} \frac{\left(\sum_{k=1}^{J} e^{u_k/\mu_2}\right)^{\mu_2/\mu_1}}{1 + \left(\sum_{k=1}^{J} e^{u_k/\mu_2}\right)^{\mu_2/\mu_1}}$$

In the multiplicative approach advocated in this paper, μ_2 depends on J. That suggests that μ_1 should depend on J as well. We derive an expression for μ_1 as a function of μ_2 by assuming that the variance of ζ_{ig} stays constant in J and using the fact that ζ_{ig} and ϵ_{ij} are distributed independently:

$$\left(\frac{\mu_2}{\mu_1}\right)^2 = \frac{var(\epsilon_{ij})}{var(\zeta_{ig} + \epsilon_{ij})} = \frac{var(\epsilon_{ij})}{var(\zeta_{ig}) + var(\epsilon_{ij})} = \frac{(\mu_2 \pi)^2/3}{var(\zeta_{ig}) + (\mu_2 \pi)^2/3}$$

$$\implies \mu_1 = \sqrt{\frac{3var(\zeta_{ig})}{\pi^2} + \mu_2^2} \tag{5}$$

A natural approach is to specify $\mu_1 = \sqrt{a + \mu_2(J)^2}$ and estimate a. The resulting Berry (1994) inversion of the share function (keeping track of μ_1) is:

$$\ln(s_j) - \ln(s_0) = \frac{u_j}{\mu_1} + \frac{\mu_1 - \mu_2}{\mu_1} \ln(s_{j|g}).$$

which again would be straightforward to estimate with non-linear techniques. Note that in this formulation, σ varies with J. This $\sigma(J)$ is not directly estimated, but can be computed with:

$$\sigma(J) = \frac{\mu_2(J)}{\mu_1(J)} = \frac{\mu_2(J)}{\sqrt{a + \mu_2(J)^2}}$$

4 Monte Carlo Results

We now turn to Monte Carlo simulations of our additive and multiplicative models. Our first goal is to see how standard logit based models perform when the data is actually generated according to one of our product congestion models. In particular, we examine how the standard models do at estimating cross-price elasticities and the welfare effects of new product introductions.

The rows of Table 1 and Table 2 contain various specifications of our additive and multiplicative nested logit models. In all specifications, we simulate data from a very large number of markets (N=1000). 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 2 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. To simplify things, price is exogenously drawn from a log-normal distribution. In all models, consumers' utility functions have a coefficient on price set at -1 and a constant of -0.5. As is standard, the utility from the outside alternative is normalized to zero.

The various specifications in the two tables differ in three dimensions. First is the type of model used to generate the data, additive (specifications (A1)-(A6), or multiplicative, (M1)-(M6)). Second is the parameter measuring product congestion in the particular model, γ or τ. We also vary σ, measuring the strength of nesting. Because of the large amount of data, the "Truth" subrows in the tables are not only the true values of these quantities, but also the estimation results from our congestion models. The "Nested Logit" subrow contains the results of naive nested logit estimation on these data.

The first row of Table 1 contains results for the pure congestion version of the additive 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

Table 1: Monte Carlo Results for Additive Model

Model		Own-Price Elasticity	Cross-Price Elasticity	Outside good P Elasticity	Welfare 2 Products	Welfare 10 Products	Percent Increase
A1-Pure Congestion	True Estimate	-1.19	0.07	0.03	0.20	0.20	0.0%
=1, =0.8	NL Estimate	-32.06	5.79	0.01	1.26	1.26	0.6%
A2	True Estimate	-1.18	0.07	0.03	0.21	0.26	26.0%
=0.95, =0.8	NL Estimate	-1.62	0.24	0.01	0.56	0.74	32.4%
A3	True Estimate	-1.17	0.08	0.04	0.23	0.41	79.0%
=0.8, =0.8	NL Estimate	-1.25	0.14	0.03	0.32	0.62	94.4%
A4	True Estimate	-1.16	0.10	0.06	0.27	0.63	133.6%
=0.5, =0.8	NL Estimate	-1.18	0.11	0.05	0.29	0.72	144.9%
A5	True Estimate	-1.82	0.18	0.03	0.21	0.24	15.5%
=0.95, =0.5	NL Estimate	-2.53	0.40	0.01	0.54	0.65	19.7%
A6	True Estimate	-4.38	0.62	0.03	0.20	0.22	5.9%
=1, =0.2	NL Estimate	-6.03	0.99	0.01	0.53	0.57	8.3%

own-price elasticity[13] to be -32.06, while the actual own-price elasticity is -1.19. Within-group cross price elasticities are also off by two orders of magnitude, and estimates of across-group (to the outside alternative) price elasticities are about 18% of their true value. The last three columns of the table show the estimated welfare effects of going from 2 to 10 products. While in actuality, there is no welfare gain to this experiment (since in a pure congestion model new products "completely" crowd out the old ones), the nested logit estimates suggest minor gains. Interestingly, in this case the nested logit model does a reasonable job at matching welfare gains, but a terrible job at price elasticities.[14]

There is a clear intuition 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 σ (e.g. in (A1), the nested logit model estimates $\sigma = 0.005$ while in truth, $\sigma = .8$). The reason for this can be seen by comparing the estimating equation for the standard nested logit model:

$$\ln(\frac{s_j}{s_0}) = X_j\beta - \alpha p_j + (1-\sigma)\ln(s_{j|g}) + \xi_j \tag{6}$$

[13]The elasticities reported in the tables are averages across the entire dataset. For example, average own-price elasticity is the average of the estimated price elasticities over all the products in the dataset. The average cross price elasticity is the average of all the cross price elasticities in the data (i.e. the average of the cross price elasticities between each product and every other product).

[14]This does match the fast food franchise story in the previous section, where the nested logit model predicts $\sigma = 0$, thus correctly measuring the welfare gains due to the entry of BK to be 0.

Table 2: Monte Carlo Results for Multiplicative Model

Model		Own-Price Elasticity	Cross-Price Elasticity	Outside good P Elasticity	Welfare 2 Products	Welfare 10 Products	Percent Increase
M1	True Estimate	-1.55	0.22	0.05	0.32	0.76	135.1%
=-0.1, J =0.8	NL Estimate	-2.01	0.35	0.14	0.36	0.89	145.0%
M2	True Estimate	-1.72	0.23	0.05	0.30	0.58	93.5%
=-0.2, J =0.8	NL Estimate	-2.40	0.51	0.14	0.41	0.86	109.7%
M3	True Estimate	-1.88	0.25	0.05	0.28	0.44	55.9%
=-0.3, J =0.8	NL Estimate	-2.96	0.75	0.14	0.49	0.88	78.5%
M4	True Estimate	-2.14	0.27	0.04	0.26	0.32	23.6%
=-0.4, J =0.8	NL Estimate	-3.98	1.16	0.14	0.60	0.92	53.8%
M5	True Estimate	-1.92	0.44	0.03	0.72	0.90	24.5%
=-0.4, J =0.5	NL Estimate	-6.53	2.28	0.18	1.51	1.91	26.2%
M6	True Estimate	-1.85	0.56	0.01	2.73	3.01	10.4%
=-0.4, J =0.2	NL Estimate	-16.92	6.63	0.22	5.10	5.55	8.8%

to the estimating equation in the additive model:

$$\ln(\frac{s_j}{s_0}) = X_j\beta - \alpha p_j + (1-\sigma)\ln(s_{j|g}) + \sigma\ln(R_j(J)) + \xi_j \tag{7}$$

Comparing the two equations, note that the estimating equation (6) has a missing variable, $\sigma\ln(R_j(J))$. 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, $\ln(s_{j|g})$, will also decline in J, so the omitted variable will be positively correlated with $\ln(s_{j|g})$ (or the typical instrument for $\ln(s_{j|g})$, i.e. J). This will tend to bias the estimate of σ downwards in the standard nested logit model. The underestimate of σ suggests too much insulation between groups. As such, across-group substitution is estimated to be too weak, and within-group substitution too strong.[15]

Models (A2) through (A6) perturb the parameters of the model. In (A2) through (A4), the congestion parameter γ is varied. As would be expected, the nested logit estimates are closer to the truth as γ decreases (recall that $\gamma = 0$ implies no congestion, i.e. the standard nested logit model **is** the truth). However, even at $\gamma = 0.5$, there are still significant biases in the nested logit results. Models (A5) and (A6) change the nesting parameter σ. While changing σ affects the absolute levels of the results, it does not appear to significantly change the percentage level of bias.

[15] For the multiplicative model, we also find that the standard nested logit model seriously underestimates the ratio μ_2/μ_1 and overestimates within group substitution. On the other hand, with the multiplicative specification, the nested logit also overestimates across-group substitution. This may be due to the fact that in the multiplicative specification, own-price elasticities are typically overestimated by more than with the additive specification.

Results for the multiplicative model, presented in Table 2 are similar. We parameterize μ_2, the scale parameter for variance within the product nest, as:

$$\mu_2 = 2\frac{J^\tau}{1 + J^\tau}$$

Following Equation 5, we specify $\mu_1 = \sqrt{a + \mu_2(J)^2}$. Under this specification, μ_2 is normalized to 1 for single product markets and $\tau = 0$ implies a standard nested logit model. We generate data for the cases of $a = 0.525$, $a = 3$, and $a = 24$ which correspond to $\mu_2/\mu_1 = 0.8$, 0.5 and 0.2 for a single product market. As τ decreases from 0, μ_2 and the ratio μ_2/μ_1 $(= \sigma)$ decrease and the market becomes more and more congested. Each row compares true and estimated results for models with successively lower values of τ.

Similar to the previous case, the standard nested logit model overestimates own- and cross- price elasticities. The difference between the two cases becomes greater as τ decreases. The estimated own-price elasticity is 30% away from the truth for M1, and 86% greater for M4. Just as striking are the welfare results. For M1, both models find large gains from going from 2 product to 10 products. However, for M4, the true model shows a 23.6% gain in welfare from adding 8 products to the market. The standard model predicts a 53.8% gain. Specifications M5 and M6 show that as σ decreases, the nested logit model does a better job of estimating welfare changes but a worse job of estimating elasticities.

For model (A2), Table 3 compares estimates of elasticities and welfare across markets with different numbers of products. Most important to note is that the standard model overestimates within-group cross-price elasticities and underestimates outside alternative price elasticities for *all* market sizes. This is likely a result of the downward bias imparted on σ described above. While the true sigma is equal to 0.8, the estimated sigma is just 0.196.

For the multiplicative model, Table 4 breaks out the $\tau = -0.4$ case by number of products. The standard model over-predicts price elasticities and, in percentage terms, predicts a much smaller change in own-price elasticity as the number of products increases. From the welfare changes, we see that the $\tau = -0.4$ case is close to a full-crowding model. There is almost no welfare gain after the 4th product. Intuitively, the standard model tries to capture this by estimating little differentiation between products (which is a very low σ) but doing so causes the model to drastically overpredict price elasticities.

In summary, these Monte-Carlo results show that if there is in fact product congestion, estimation by

Table 3: Monte Carlo Results for Additive Model (A1)

Num of Products			Own-Price Elasticity		Cross-Price Elasticity		Outside Option Price Elasticity		Welfare	
	Truth	Estimate	Truth	Estimate	Truth	Estimate	Truth	Estimate	Truth	Estimate
2	0.50	0.31	-1.646	-1.703	0.055	0.085	0.740	0.554	0.195	0.300
3	0.50	0.31	-1.918	-1.901	0.038	0.058	0.448	0.339	0.209	0.336
4	0.50	0.31	-2.034	-1.986	0.030	0.046	0.330	0.251	0.222	0.364
5	0.50	0.31	-2.123	-2.058	0.025	0.038	0.279	0.213	0.234	0.387
6	0.50	0.31	-2.191	-2.112	0.022	0.033	0.215	0.165	0.245	0.406
7	0.50	0.31	-2.198	-2.112	0.019	0.029	0.173	0.134	0.255	0.424
8	0.50	0.31	-2.221	-2.130	0.017	0.026	0.161	0.125	0.265	0.439
9	0.50	0.31	-2.232	-2.137	0.016	0.024	0.146	0.113	0.275	0.453
10	0.50	0.31	-2.275	-2.175	0.014	0.022	0.130	0.102	0.284	0.467

Table 4: Monte Carlo Results for Multiplicative Model (M4)

Num of Products	Ratio		Own-Price Elasticity		Cross-Price Elasticity		Outside Option Price Elasticity		Welfare	
	Truth	Estimate	Truth	Estimate	Truth	Estimate	Truth	Estimate	Truth	Estimate
2	0.80	0.32	-1.53	-3.17	0.40	1.96	0.16	0.23	0.26	0.60
3	0.75	0.32	-1.81	-3.81	0.32	1.32	0.12	0.17	0.29	0.67
4	0.72	0.32	-2.01	-4.14	0.27	1.00	0.10	0.13	0.30	0.72
5	0.70	0.32	-2.18	-4.33	0.24	0.81	0.08	0.11	0.31	0.77
6	0.68	0.32	-2.32	-4.46	0.22	0.68	0.07	0.10	0.31	0.81
7	0.66	0.32	-2.45	-4.56	0.20	0.58	0.06	0.09	0.32	0.84
8	0.64	0.32	-2.56	-4.63	0.18	0.51	0.06	0.08	0.32	0.87
9	0.63	0.32	-2.67	-4.68	0.17	0.46	0.05	0.07	0.32	0.90
10	0.62	0.32	-2.76	-4.73	0.16	0.41	0.05	0.06	0.32	0.92

standard methods can give biased and very misleading estimates. These biases can be up to an order of magnitude.

4.1 Other Types of Congestion

A caveat of the above monte-carlo results is that the simulated data comes from exactly the congestion process we specify. Here we briefly examine how our models perform when congestion comes from some alternative model. Since our models are misspecified in this case, we don't expect to recover parameters of interest exactly, but we do expect to perform better than models with standard logit errors. The data used for estimation in Table **??** are generated by a random coefficients (on observable characteristics) model. There are random coefficients on both the constant term and on a single observed characteristic that is distributed uniformly across firms. Congestion in unobserved characteristic space is generated by a one dimensional locational (with transport costs) model. Specifically, products differ in their location in a Hotelling linear city model. Products spread equally across the linear city. Thus, markets with more products have more

Table 5: Monte Carlo Results for Locational Congestion

	Low Transport Costs			Medium Transport Costs			High Transport Costs		
Num. of Products	True Elasticities	RCM	RCM + Mult	True Elasticities	RCM	RCM + Mult	True Elasticities	RCM	RCM + Mult
2	2.73	2.87	2.85	1.28	2.47	1.55	0.42	1.77	0.61
3	4.16	4.28	4.22	2.75	3.57	2.51	1.00	2.41	1.03
4	5.05	5.13	5.02	3.74	4.21	3.67	1.83	2.77	1.97
5	5.65	5.71	5.62	4.46	4.63	4.20	2.63	2.99	2.39
6	6.09	6.12	6.07	5.00	4.92	4.69	3.28	3.14	2.92
7	6.42	6.43	6.38	5.42	5.13	5.01	3.80	3.26	3.24
8	6.68	6.67	6.63	5.76	5.30	5.30	4.24	3.34	3.58
9	6.89	6.86	6.87	6.04	5.43	5.54	4.60	3.41	3.81
10	7.06	7.01	7.09	6.27	5.53	5.77	4.91	3.46	4.06

congestion in unobserved characteristic space.[16]

Table **??** shows estimates of own-price elasticities for three different data sets. The data sets differ in the magnitude of transportation costs in the linear city. As transport costs increase, the importance of these unobserved product characteristics increases relative to the importance of the observed product characteristics. As a result, one can interpret the different data sets as capturing differing levels of success of the econometrician in measuring relevant product characteristics in the market of interest.

For each data set, three sets of elasticities are reported. In the first column are the true elasticities generated by the model. The second column are elasticities derived from estimating a standard random coefficients model (plus logit errors). The third column are estimates from a standard random coefficients model plus our multiplicative adjustment.[17] With the lowest transportation costs, the misspecification of unobserved product differentiation does not cause significant bias in the price elasticities. Both the standard RCM model and the congestion model do a reasonable job. With medium transport costs, the accuracy of the RCM results decreases - while the true elasticities range from 1.28 (in a market with two products) to

[16] The outside good is assumed to incur no transport costs. We also include very low variance logit errors in the data generating process to prevent zero market shares (to generate a small variance (relative to other consumer heterogeneity) logit error, we inflated the means and variances of the random coefficients - these were $\beta_{0i}\tilde{}N(0,5)$, $\beta_{1i}\tilde{}N(5,5)$). As a result, unobserved product characteristic space includes both the congestable linear dimension and a small, non-congestable logit error dimension.

[17] The multiplicative model worked a bit better than the additive one on this Hotelling style unobserved product differentiation. Note also that while welfare calculations were more accurate with our multiplicative and additive models than the standard random coefficient model, neither model obtains particularly realistic welfare numbers. This is to be expected, as the top of the demand curve is going to be highly dependent on the form of unobserved product differentiation.

6.27 (in a market with 10 products), the RCM estimates range from 2.47 to 5.53. Note that the upward bias in elasticities in small markets and the downward bias in large markets corresponds to some of the intuition developed in the introduction. Standard logit errors are unable to fully capture the fact that through congestion, elasticities increase in crowded markets. In contrast, our congestion model performs significantly better - elasticities range from 1.55 to 5.77. In the last group of results, the biases in the standard RCM model are even larger, while our congestion model still performs well. For small markets, for example, where the true elasticity is 0.42, the RCM model estimates an elasticity of 1.77. Our congestion model estimates it to be 0.61. In summary, while it is hard to address potential misspecification issues (as there is a continuum of potential misspecifications), these results support our intuition, suggesting that our congestion models can do significantly better than standard logit based models at addressing arbitrary congestion in unobserved product characteristic space.

5 Example

Rysman (2002) studies a data set on the Yellow Pages industry, measuring the positive feedback loop between consumers' choice of directory to use (which is driven by the amount of advertising in the directory) and retailer's 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 2 shows the percentage of consumers served by different numbers of directories. The variance in this number of directories makes this is a natural place to apply the techniques presented in this paper. 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 paper.[18] 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.

The data set consists of observations on the number of uses, per household, per month, in the distribution

[18]The policy question is whether or not welfare improves as competition increases. Multiple directories reduce market power but dissipate network effects.

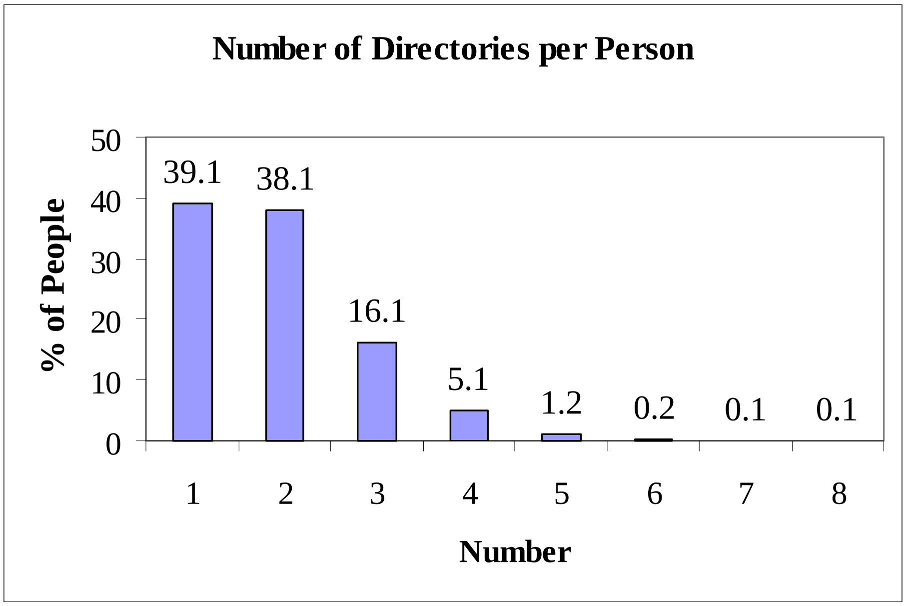

Figure 2: Percentage of People Served by Each Number of Directories

areas of 428 directories in 1996.[19] We assume a representative consumer needs information of the kind they could find in the Yellow Pages M times per month . The exogenous parameter M is constant across markets. Each time a consumer needs information, the consumer 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:

$$u_{ij} = \beta_1 \ln(A_j) + X_j\beta_2 + \xi_j + \varepsilon_{ij}$$

The variable A_j is the quantity of advertising at directory j and the matrix X_j represents demographic variables that may affect usage.[20] The variable ξ_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 basic model in 3 different ways: with standard logit errors, with the additive adjustment,

[19] The data was collected by National Yellow Pages Monitor for use by Yellow Pages publishers and advertising agencies. NYPM survey respondents maintain diaries of their Yellow Pages usage for 1 week. NYPM normally surveys between 1,000 and 3,000 people per MSA, although NYPM used 11,200 respondents in the Los Angeles area. This usually results in at least a few hundred respondents even for very small directories.

[20] As a measure of advertising, Rysman uses the number of pages in a book times the number of columns in a directory. The is number is multiplied by 0.8 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.

and with the multiplicative adjustment. To see clearest what the data tells us about congestion, we use non-parametric specifications of the additive and multiplicative terms. That is, we allow the number of retailers or the variance scale parameter to take on different values for each number of products in the market. 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 1 or more competitors for another group of consumers. Furthermore, we observe distribution areas but we cannot distinguish how much usage comes from different portions of a directory's distribution area.

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[21]) in directory j's total market, and sub-markets (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 sub-market. 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 additive model, we simply take the additive term to be the population weighted average of R_j across submarkets. In that case, we estimate:

$$\begin{aligned} \ln(s_j) - \ln(s_0) &= \alpha \ln(A_j) + X_j\beta + R_j + \xi_j \\ \text{where } R_j &= \sum_{k \in K(j)} \psi_{jk} R_{J(k)} \end{aligned}$$

Here, $K(j)$ is the set of sub-markets in j's market area, ψ_{jk} is the percentage of j's population that lives in sub-market k. The parameter R_J is to be estimated, separately for each J and $J(k)$ is the number of products in sub-market k.

To implement the multiplicative model, we push the model to its logical extreme and assume that the scale parameter μ differs for directories across sub-markets. That is, the variance of ε_{ij} differs for the same product based on the number of competitors for consumer i. Therefore, the market share for product j is:

$$s_j = \sum_{k \in K(j)} \psi_{jk} \frac{\exp\left(u_j/\mu_{J(k)}\right)}{\sum_{i \in D(k)} \exp\left(u_i/\mu_{J(k)}\right)}$$

[21]We assume that $M = 26$. The highest number of uses per household in our data set is 23.6 with an average of 11.4. The average for s_0 in our data set is 47.7%.

where $D(k)$ is the set of directories in sub-market k and μ_J is to be estimated separately for each J. The variable u_j is the mean utility for product j. For a given set of parameters μ, we can infer (via a fixed point algorithm) the vector of mean utilities u that implies sub-market shares that aggregate up to the market shares we observe. Then we can estimate the remaining parameters via the equation:

$$u_j = \alpha \ln(A_j) + X_j \beta + \xi_j$$

We estimate all 3 specifications by the Generalized Method of Moments (Hansen (1982)) using the same set of instruments as in Rysman (2002).

We observe very few markets with more then 5 directories so, in practice, we restrict markets with 6, 7 or 8 directories to have the same adjustment parameter. Results appear in Table 6. Parameter estimates show that the additive specification and the multiplicative specification produce very similar results. Crowding appears to be important in both models. The parameters for the additive adjustment are close to being monotonic in J and decrease at a decreasing rate. The parameters for the multiplicative model show that the variance for markets with multiple directories are much smaller then for those with only one directory. The parameters do not vary much in markets with more than one directory, suggesting that this model could be estimated with a single μ for all oligopoly markets. The biggest change in the explanatory variables across the 3 models is that the coefficient on advertising is lower in the multiplicative model. The low coefficient compensates for the reduced variance in crowded markets.

Table 7 presents summary statistics. The first column presents the elasticity of usage from advertising. As in our monte-carlo results, the standard logit model overestimates (advertising) elasticities. In single product markets the standard logit overestimates the advertising elasticity by 29% relative to the additive model and 76% relative to the multiplicative model. Another feature to notice is how the crowding models generate larger increases in elasticity as the number of products increase. When the number of products goes from 1 to 8, the standard logit model shows that elasticity increases by 18% whereas the additive model finds that elasticity increases by 30% and the multiplicative model finds 79%. 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.

Equally as striking are the welfare calculations. The logit model predicts that even the 7th and 8th Yellow Pages directories imply non-trivial welfare increases, over a third of what the first directory generates. On the

Table 6: Estimation Results for Yellow Pages Data

	Standard		Additive		Multiplicative	
Variable	Coef	Std Err	Coef	Std Err	Coef	Std Err
advertising	0.75	(0.08)	0.64	(0.07)	0.22	(0.04)
constant	-6.55	(1.21)	-5.11	(1.01)	-2.08	(0.43)
% urban population	-0.01	(0.01)	-0.01	(0.00)	0.00	(0.00)
% lived in diff county	0.08	(0.02)	0.07	(0.01)	0.02	(0.01)
% lived in diff state	0.05	(0.03)	0.03	(0.02)	0.01	(0.01)
% own house	-0.01	(0.01)	-0.02	(0.01)	-0.01	(0.00)
% grad hi school	-0.05	(0.02)	-0.04	(0.01)	-0.01	(0.00)
% grad college	-0.02	(0.02)	0.00	(0.02)	-0.01	(0.01)
per cap income	0.03	(0.03)	0.01	(0.02)	0.01	(0.01)
telco book	1.09	(0.13)	1.05	(0.10)	0.38	(0.05)
county pop. growth rate	0.01	(0.02)	0.01	(0.01)	0.01	(0.01)
% take public trans.	-0.05	(0.04)	-0.05	(0.03)	-0.01	(0.01)
% have not moved	0.07	(0.02)	0.06	(0.02)	0.02	(0.01)
pop. density	-1.3E-04	(6.1E-05)	-6.9E-05	(3.5E-05)	-3.0E-05	(1.3E-05)
Adjustment J=1			0.00	Fixed	1.00	Fixed
J=2			-0.35	(1.00)	0.35	(0.03)
J=3			-0.28	(0.99)	0.36	(0.04)
J=4			-0.71	(0.98)	0.30	(0.03)
J=5			-0.80	(1.00)	0.32	(0.04)
J=6, 7, 8			-0.91	(0.99)	0.34	(0.05)

other hand, the additive and multiplicative specifications imply 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 additive model, welfare increases by 145% and under the multiplicative model, welfare increases by 109%. Note that the additive and multiplicative models actually find that welfare decreases for some increases in the choice set. This result would likely disappear if we put more structure on our additive and multiplicative J functions.

6 Conclusion

This paper 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 product space can lead to significantly biased estimates of elasticities and welfare changes. We suggest two solutions, an additive and a multiplicative adjustment to the standard estimating equations. We present structural interpretations of

Table 7: Summary Variables for Yellow Pages Data

		Elasticities			Welfare		
		Standard	Add	Mult	Standard	Add	Mult
Directories	1	0.58	0.45	0.33	0.20	0.27	0.25
	2	0.61	0.52	0.53	0.36	0.36	0.20
	3	0.63	0.54	0.53	0.51	0.50	0.28
	4	0.65	0.57	0.65	0.63	0.46	0.22
	5	0.66	0.58	0.61	0.74	0.50	0.33
	6	0.67	0.58	0.58	0.84	0.53	0.43
	7	0.68	0.59	0.58	0.93	0.60	0.48
	8	0.68	0.59	0.59	1.02	0.66	0.53

our solutions, showing how they could arise from the appropriate agent maximization problem. We present Monte Carlo evidence that shows the efficacy of our adjustments, and we examine how our adjustments perform in a real data set.

An interesting question is what circumstances are appropriate for which adjustment. The additive adjustment is typically easier to implement then the multiplicative adjustment. It can specified in a linear manner, and can easily be extended to multi-nested models or random coefficient frameworks. While the multiplicative model can be applied in those circumstances, one must maintain that each choice has the same variance or abandon the random utility interpretation of the model. Conversely, the multiplicative model can be applied even in the simple logit case where the researcher is not willing to specify an "outside option". While the two models seem to obtain similar results, they are not identical, so the choice of model might be important for specific applications. In this case, it might be fruitful to do formal non-nested testing of the models. Lastly, note that it is also possible to combine the two models - i.e. include *both* additive and multiplicative adjustments in the estimating equation.

References

[1] Anderson, S., A. de Palma and J-F Thisse. (1992), *Discrete Choice Theory of Product Differentiation*, Cambridge: The MIT Press.

[2] Arcidiacono, P. (1999) "Option Values, College Quality, and Earnings: A Dynamic Model College and Major Choice" mimeo, Duke University.

[3] Benkard, C.L. and Bajari, P. (2001) "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach", mimeo, Stanford University

[4] Berry, S.T. (1994) "Estimating Discrete Choice Models of Product Differentiation." *Rand Journal of Economics*: 25, pp 242-262.

[5] Berry, S.T. and A. Pakes. (1999) "Estimating the Pure Hedonic Discrete Choice Model." mimeo, Yale University.

[6] Berry, S.T.; J. A. Levinsohn and A Pakes. (1995) "Automobile Prices in Market Equilibrium." *Econometrica*: 63(4), pp. 841-890.

[7] Berry, S.T. and J. Waldfogel. (1999) "Free Entry and Social Inefficiency in Radio Broadcasting," *RAND Journal of Economics:* 30(3), pp 397-420.

[8] Bresnahan, T.F. (1987) "Competition and Collusion in the American Auto Industry: The 1955 Price War." *Journal of Industrial Economics*: 35, pp. 457-482.

[9] Bresnahan, T.F.; S. Stern and M. Trajtenberg. (1997) "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s." *Rand Journal of Economics*: 28(0), pp S17-S44.

[10] Cardell, N. S. (1997) "Variance Components Structures for the Extreme Value of Logistic Distributions with Application to Models of Heterogeneity." *Econometric Theory*: 113(2), pp. 185-213.

[11] Feenstra R.C. and J.A. Levinsohn (1995) "Estimating Markups and Market Conduct with Multidimensional Product Attributes." *Review of Economic Studies*: 62(1), pp. 19-52.

[12] McFadden, D. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), Frontiers in Econometrics, 105-142, Academic Press: New York.

[13] Nevo, A. (2001) "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*: 69(2), 307-342.

[14] Petrin, A. (1999) "Quantifying the Effects of New Products: The Case of the Minivan." forthcoming *Journal of Political Economy*.

[15] Rysman, M. (2002) "Competition Between Networks: A Study of the Market for Yellow Pages." mimeo, Boston University.

[16] Shum, M. (1999) "The Effects of Advertising in Markets with Brand Loyalty: the Case of Breakfast Cereals." mimeo, University of Toronto.

[17] Trajtenberg, M. (1990) *Economic Analysis of Product Innovation,* Cambridge: Harvard University Press.