Supplement to
Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects

Daniel A. Ackerberg
UCLA

Marc Rysman
Boston University

June 20, 2006

1 Introduction

This document is a supplement to Ackerberg and Rysman (2005). In that paper, we discuss problems with standard techniques for discrete choice estimation in the context of consumers that face different numbers of products. We propose a solution based on adding a function of the number of products to the estimation equation, which we motivate with a model of retail conjestion. Section 6 of the main paper introduces a second solution that allows the variance of the logit error to be smaller in markets with more products. This solution entails a multiplicative adjustment rather than an additive adjustment. Here, we elaborate on the multiplicative solution. We show how this feature can arise from a model in which products in crowded markets differentiate into dimensions that consumers care less about. We provide Monte Carlo analysis and an empirical example.

2 The Multiplicative Model

As in the paper, we present our ideas in the context of the nested logit model. Let utility to consumer i from buying product j be $u_{ij} = \beta_0 + X_j\beta_1 + \zeta_{ig} + \varepsilon_{ij}$ where

X_j are observable characteristics, ζ_{ig} is a consumer idiosyncratic preference for the group of products g and ε_{ij} is the consumer idiosyncratic preference for the product j. We assume that ε_{ij} is distributed Extreme Value with variance scale parameter μ_2 and ζ_{ig} is distributed such that $\zeta_{ig} + \varepsilon_{ij}$ is distributed Extreme Value with variance scale parameter μ_1. Naturally, $\mu_1 \geq \mu_2$. We sometimes denote $\sigma = \mu_2/\mu_1$.

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 $\mu_2 = \mu_2(J; \tau)$.[1] 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 described in the main paper (Equations 2 and 3). Relative to a constant μ_2,finding that $\mu_2(J; \tau)$ decreases in J leads us to find that the elasticity to price is higher in more crowded markets and that additional products provide lower welfare benefits.

2.1 A Structural Interpretation

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

[1] If μ_2 depends on J, then μ_1 (and σ) does also. We address this issue in Section 2.2

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.

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.[2] 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.

[2] 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.

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 (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 the main paper 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
$$z_j^l = \begin{cases} b & \textit{if } l = j, \quad j, l = 1 \ldots J \\ -b & \textit{otherwise} \end{cases}$$
$$z_0^l = -b \quad l = 1, \ldots, J.$$

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 is satisfied if:

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

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

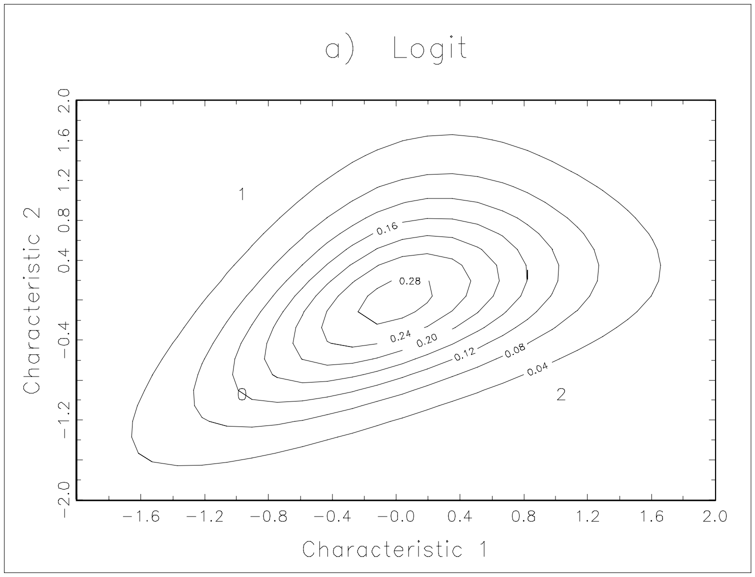

Figure 1: Consumer Distribution in the Address Model that Matches a Logit Model

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 1, 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 1 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 $I : [1, J!] \times [1, J] \rightarrow [1, J]$ 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:

$$s_j = \frac{\exp(u_j/\mu(J))}{1+\sum_{k=1}^{J}\exp(u_k/\mu(J))} \tag{2}$$

2.2 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. 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:

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

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

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:

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

3 Monte Carlo Results

We now turn to Monte Carlo simulations of the multiplicative model. Our first goal is to see how standard logit based models perform when the data is actually generated according to our product congestion model. In particular, we examine how the standard models do at estimating cross-price elasticities and the welfare effects of new product introductions when data is actually generated from the multiplicative model. A second set of results generates data from an alternative crowding model and checks whether the multiplicative model is superior.

3.1 Multiplicative congestion

The rows of Table 1 contain various specifications of our 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

Table 1: Monte Carlo Results for Multiplicative Model

Parameters	Estimate	Own-Price Elasticity	Cross-Price Elasticity	Outside good P Elasticity	Welfare 2 Products	Welfare 10 Products	Percent Increase
M1	Truth	-1.55	0.22	0.05	0.32	0.76	135.1%
τ=-0.1, σ(J=1)=0.8	NL Estimate	-2.01	0.35	0.14	0.36	0.89	145.0%
M2	Truth	-1.72	0.23	0.05	0.30	0.58	93.5%
τ=-0.2, σ(J=1)=0.8	NL Estimate	-2.40	0.51	0.14	0.41	0.86	109.7%
M3	Truth	-1.88	0.25	0.05	0.28	0.44	55.9%
τ=-0.3, σ(J=1)=0.8	NL Estimate	-2.96	0.75	0.14	0.49	0.88	78.5%
M4	Truth	-2.14	0.27	0.04	0.26	0.32	23.6%
τ=-0.4, σ(J=1)=0.8	NL Estimate	-3.98	1.16	0.14	0.60	0.92	53.8%
M5	Truth	-1.92	0.44	0.03	0.72	0.90	24.5%
τ=-0.4, σ(J=1)=0.5	NL Estimate	-6.53	2.28	0.18	1.51	1.91	26.2%
M6	Truth	-1.85	0.56	0.01	2.73	3.01	10.4%
τ=-0.4, σ(J=1)=0.2	NL Estimate	-16.92	6.63	0.22	5.10	5.55	8.8%

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 differ in two dimensions. First is the parameter measuring product congestion in the particular model, τ. The second is the parameter measuring the strength of nesting σ. Because of the large amount of data, the "Truth" subrows in the table 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.

Results for the multiplicative model, presented in Table 1 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 3, 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 additive 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.[3]

For the multiplicative model, Table 2 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 standard methods can give biased and very misleading estimates. These biases can be up to an order of magnitude.

[3] For the multiplicative model, we find that the standard nested logit model overestimates across-group substitution, unlike for the additive model. 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.

Table 2: Monte Carlo Results for Multiplicative Model (M4)

Num of Products	Ratio μ2/μ1 (=σ)		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

3.2 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 model performs when congestion comes from some alternative model. Since our model is 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 3 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 congestion in unobserved characteristic space.[4]

Table 3 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

[4]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 $\beta_{0i} \sim N(0,5)$, $\beta_{1i} \sim N(5,5)$). As a result, unobserved product characteristic space includes both the congestable linear dimension and a small, non-congestable logit error dimension.

Table 3: Monte Carlo Results for Locational Congestion

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

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.[5] 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 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

[5]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.

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.

4 Example

Here, we use the data in Rysman (2004) to study the role of our adjustments. This section differs from the main paper in that it estimates both the additive and the multipicative model. 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:

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

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 three specifications by the Generalized Method of Moments using the same set of instruments as in Rysman (2004).

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 4. 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 5 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 22% relative to the additive model and 83% 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 16% whereas the additive model finds that elasticity increases by 29% and the multiplicative model finds 164%. 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 other hand, the additive and multi-

Table 4: Estimation Results for Yellow Pages Data

	Standard		Additive		Multiplicative	
Variable	Coef	Std Err	Coef	Std Err	Coef	Std Err
advertising	0.71	(0.07)	0.63	(0.07)	0.22	(0.04)
constant	-6.08	(1.07)	-4.94	1.01	-2.08	(0.43)
% urban population	-0.23	(0.01)	-0.01	(0.00)	0.00	(0.00)
% lived in diff county	0.08	(0.02)	0.06	(0.01)	0.02	(0.01)
% lived in diff state	0.05	(0.02)	0.03	(0.02)	0.01	(0.01)
% own house	-0.02	(0.01)	-0.21	(0.01)	-0.01	(0.00)
% grad hi school	-0.04	(0.01)	-0.04	(0.01)	-0.01	(0.00)
% grad college	-0.02	(0.02)	-0.01	(0.02)	-0.01	(0.01)
per cap income	0.03	(0.02)	0.02	(0.02)	0.01	(0.01)
telco book	1.16	(0.10)	1.02	(0.10)	0.38	(0.05)
county pop. growth rate	0.00	(0.02)	0.02	(0.01)	0.01	(0.01)
% take public trans.	-0.04	(0.03)	-0.04	(0.03)	-0.01	(0.01)
% have not moved	0.07	(0.02)	0.05	(0.02)	0.02	(0.01)
pop. density	-1.1E-04	(3.9E-05)	-7.2E-05	(3.5E-05)	-3.0E-05	(1.3E-05)
Adjustment J=1			0.00	Fixed	1.00	Fixed
J=2			-0.35	(0.14)	0.35	(0.03)
J=3			-0.34	(0.18)	0.36	(0.04)
J=4			-0.74	(0.22)	0.30	(0.03)
J=5			-0.87	(0.31)	0.32	(0.04)
J=6, 7, 8			-0.97	(0.36)	0.34	(0.05)

plicative 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 changes little with the number of directories, or even decreases. Note that both the additive and multiplicative models 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.

Table 5: Summary Variables for Yellow Pages Data

Directories	Elasticities Standard	Elasticities Add	Elasticities Mult	Welfare Standard	Welfare Add	Welfare Mult
1	0.55	0.45	0.30	0.20	0.27	0.24
2	0.58	0.52	0.41	0.36	0.36	0.27
3	0.60	0.55	0.50	0.51	0.50	0.26
4	0.61	0.56	0.57	0.63	0.46	0.25
5	0.62	0.57	0.63	0.74	0.50	0.59
6	0.63	0.57	0.69	0.84	0.53	0.22
7	0.64	0.58	0.74	0.93	0.60	0.20
8	0.64	0.58	0.79	1.02	0.66	0.19

5 Conclusion

This supplement and the associated 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 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] Ackerberg, D. A., and Rysman, M., (2005) "Unobservable Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects," *RAND Jouranl of Economics, 36*, 771-788.

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

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

[4] Rysman, M. (2004) "Competition Between Networks: A Study of the Market for Yellow Pages," *Review of Economic Studies*: 71, 483-512.