# Estimation of Discrete Choice Models with Many Alternatives Using Random Subsets of the Full Choice Set: With an Application to Demand for Frozen Pizza

by

Michael P. Keane
University of Oxford, Nuffield College and Department of Economics

Nada Wasi
University of Michigan

October 3, 2012

**Abstract**: A common problem in estimation of discrete choice models is that the complete choice set is very large. A good example is supermarket consumer goods, like breakfast cereal, where there are often a hundred or more varieties (SKUs or UPCs) to choose from. In that case, estimation of complex discrete choice models where choice probabilities have no closed form can be very computationally burdensome. We show how use of random subsets of the full choice set can be a useful device to reduce computational burden. We apply this approach to estimating demand for frozen pizza, where there are nearly 100 varieties to choose from. We provide some interesting new results on how price changes for a particular variety of a brand lead to variety switching within the brand vs. brand switching. In particular, when a variety raises its price, most switching is to other brands, rather than to other varieties of the same brand.

## I. Introduction

In many situations, consumers confront choice problems that involve a very large number of alternatives. Examples range widely from choice of homes or residential location, choice of colleges and majors, choice of TV shows or movies, choice of occupation at a detailed level, or the choice of a breakfast cereal or a frozen food variety in a supermarket. Large choice sets can create substantial computational problems for researchers interested in estimating modern discrete choice models that allow for a rich structure of consumer taste heterogeneity.

Choice probabilities in discrete choice models with unobserved heterogeneity take the form of integrals whose dimension is comparable to the size of the choice set, and that usually have no closed form. In recent years, simulation methods have made estimation of such models feasible.[1] Nevertheless, given very large choice sets, estimation of discrete choice models can still be very computationally burdensome, particularly given the very large sample sizes often available in modern datasets on consumer behavior.

In this paper we show how the use of random subsets of the full choice set can be a useful device to reduce computational burden in such contexts. In a classic paper, McFadden (1978) showed that the use of random subsets of the full choice set has no effect on the consistency of parameter estimates in the multinomial logit model (MNL).[2] However, as we discuss below, McFadden's result does not go through in models with unobserved consumer taste heterogeneity.

Nevertheless, we present a Monte Carlo analysis that shows how use of random subsets of the full choice set leads to a negligible bias in estimates of three important discrete choice models with heterogeneity. The three models we consider are the mixed logit (MIXL) with normal mixing (N-MIXL), the mixed logit with discrete mixture-of-normals mixing (MM-MNL) and the generalized multinomial logit (G-MNL).

We also present an application to scanner data on frozen pizza. There are nearly 100 options (brands/varieties) in the full choice set. We obtain results using random subsets of 20 to 40 options, and show it makes little difference to the estimates. Substantively, we use the models to decompose price elasticities into brand/variety specific components. We find that, if a variety raises its price, most switching is to other brands, rather than other varieties of the same brand.

---

[1] Some key references are McFadden (1989), Pakes (1986), Keane (1994), McCulloch and Rossi (1994), Geweke and Keane (2001).

[2] Given the great speed of modern computers, it is now feasible to estimate MNL even with very large choice sets and very large datasets. This is because MNL generates closed form expressions for the choice probability integrals.

Since we model consumer choice at the variety level, rather than the brand level, our work is obviously related to the relatively small but growing literature on UPC or SKU-level modeling. Pioneering papers in this literature are Fader and Hardie (1996) and Andrews and Manrai (1999). As Fader and Hardie (1996) note, "in contrast to most choice models presented in the marketing literature, brand choice is rarely a final decision by itself; rather, SKU choice is a more fitting description of the overall decision process … Nevertheless, most choice modelers … assume away most of these critical details."

In our view there are two reasons that SKU-level modeling has been uncommon. The first problem, stressed by Fader-Hardie and Andrews-Manrai, is that the very large SKU-level choice sets lead to a proliferation of parameters, particularly if one allows for SKU-specific preferences. Fader and Hardie (1996) dealt with this by assuming consumers have preferences over a relatively small set of common product attributes, not over SKUs themselves. Andrews and Manrai (1999) extended this by projecting latent consumer preference vectors onto a space of lower dimension than the number of consumer types, thus conserving parameters. Other papers that adopt/extend these ideas (e.g., by incorporating state dependence in tastes for attributes into SKU-level models), are Ho and Chong (2003), Dube (2004), Chintagunta and Dube (2005), Singh, Hansen and Gupta (2005) and Inman, Park and Sinha (2008).

A second reason that SKU-level modeling has been uncommon is that, with large choice sets, computation of the choice probability integrals can be very burdensome (even if one has dealt with the proliferation of parameters problem). For this reason, all the papers cited above use simple MNL or latent class MNL models. In each case, one obtains closed form expressions for the choice probabilities. But the method of using random subsets of the full choice set makes it feasible to introduce much richer heterogeneity structures into models with large choice sets.

Thus, we offer no new approach for reducing the number parameters in SKU-level models; indeed, we use the same approach as in Fader-Hardie and Andrews-Manrai (i.e., assuming consumers care about common rather than SKU-specific attributes, and imposing a factor structure on the heterogeneity distribution, as in Elrod and Keane (1995)). But the reduced choice set method we explore here can be applied in many different types of SKU-level models, regardless of how one reduces the dimension of the parameter vector. Thus, it should be relevant for the SKU-level modeling approaches developed in all the papers cited above. We illustrate the efficacy of the reduced choice set approach using several alternative models of heterogeneity.

The outline of the paper is as follows: In Section II we describe alternative models of heterogeneity. Section III discusses the McFadden (1978) result for MNL and explains why it does not go through with heterogeneity. Section IV presents the Monte Carlo results, and gives an intuition for why the McFadden (1978) holds approximately even with heterogeneity. Section V presents the application to frozen pizza, and Section VI concludes.

## II. Alternative Models of Heterogeneity in Consumer Choice Behavior

The MNL model of McFadden (1974) was the primary basis for analysis of multinomial choice for many years. MNL assumes consumers have homogeneous tastes for observed product attributes, and that the random (unobserved) part of utility is iid. These two strong assumptions rule out persistent heterogeneity in tastes for observed and unobserved product attributes. And they imply some unrealistic properties, like the independence of irrelevant alternatives (IIA).

A number of alternative models that extend MNL to allow for taste heterogeneity have been proposed. We examine several of the most important here. All models we consider can be written in the following form: The utility to person $n$ from choosing alternative $j$ on purchase occasion (or choice scenario) $t$ is given by:

$$U_{njt} = \beta_n x_{njt} + \varepsilon_{njt} \qquad n=1,...,N; \quad j=1,...,J; \quad t=1,...,T, \tag{1}$$

where $x_{njt}$ is a $K$-vector of observed attributes of alternative $j$ facing person $n$ on occasion $t$, while $\beta_n$ is a vector of person $n$ specific utility weights on these attributes. These random coefficients capture heterogeneity in tastes for observed attributes. However, the $x_{njt}$ for $j = 1,...,J$ may also be specified to include alternative specific constants (ASCs). Then, the random coefficients on these constants capture heterogeneity in tastes for the unobserved attributes of each alternative.

The $\varepsilon_{njt}$ in (1) are random error components that capture "idiosyncratic" tastes for alternative $j$ by person $n$ at time $t$. By assuming they are distributed *iid* extreme value, we obtain the mixed logit (MIXL) family of models. As noted by McFadden and Train (2000), given proper choice of the mixing distribution the MIXL family nests (or can approximate) all random utility models. For example, if $\beta_n$ is multivariate normal, and if the variance of the $\beta_n$ vector grows large relative to that of the $\varepsilon_{njt}$ vector, the model approximates multinomial probit.

Of course, we cannot consider all possible MIXL models, so we limit ourselves to six that have been particularly important in the literature. The models are distinguished by different

specifications of the mixing distribution ($\beta_n$). The first is MNL itself, where $\beta_n = \beta$ for all $n$. Second is N-MIXL, where $\beta_n$ is distributed multivariate normal N(0,$\Sigma$) in the population. This model can approximate MNP if ASCs are included, and it is very popular in applications.

Third is the latent class (LC) model. This assumes there are $S$ discrete segments of consumers, $s=1,\ldots,S$. Each segment has its own $\beta$ vector ($\beta_s$), but there is no heterogeneity within segments. That is, $\beta_n = \beta_s \forall n \in s$. Segments are latent and $S$ is not known *a priori*.

Next are the S-MNL and G-MNL models proposed in Fiebig et al (2010). S-MNL assumes $\beta_n = \sigma_n \beta$ where $\sigma_n$ is a positive scalar that shifts the whole $\beta$ vector up or down. The motivation for this specification is work by Louviere et al (1999, 2002) and Meyer and Louviere (2007) that suggests much of the heterogeneity in SP data takes the form of scale heterogeneity. We assume $\sigma_n$ has a lognormal distribution, $\ln(\sigma_n) \sim N(\bar{\sigma}, \tau^2)$. This assures that $\sigma_n$ is positive. We normalize E($\sigma_n$)=1 for identification (i.e., we estimate only $\beta$ and $\tau$ and *calibrate* $\bar{\sigma}$ so that E($\sigma_n$)=1). Thus $\beta$ is interpretable as the mean vector of the random preference weights $\beta_n$.

The G-MNL model of Fiebig et al (2010) nests S-MNL and N-MIXL. In G-MNL, the utility to person $n$ from alternative $j$ on purchase occasion (or in choice scenario) $t$ is given by:

$$U_{njt} = [\sigma_n \beta + \gamma \eta_n + (1-\gamma)\sigma_n \eta_n]x_{njt} + \varepsilon_{njt} \tag{2}$$

where $\gamma$ is a parameter that determines how scale affects $\eta_n$. This is a special case of (1) where:

$$\beta_n = [\sigma_n \beta + \gamma \eta_n + (1-\gamma)\sigma_n \eta_n] \tag{3}$$

and $\eta_n$ is a N(0,$\Sigma$) random vector. To obtain N-MIXL one sets the scale parameter $\sigma_n=\sigma=1$. To obtain S-MNL one sets Trace($\eta_n$)=0, so the variance-covariance matrix of $\eta_n$ is degenerate.

In the general case where both scale heterogeneity (Var($\sigma_n$)>0) and normally distributed random coefficients are present (Trace($\eta_n$)>0), the parameter $\gamma$ governs how the scale of the normal errors ($\eta_n$) varies with $\sigma_n$. For instance, if $\gamma = 1$ then $\beta_n = [\sigma_n \beta + \eta_n]$ and the scale of the normal errors does not vary with that of the $\beta$ vector. But if $\gamma = 0$ then $\beta_n = [\sigma_n \beta + \sigma_n \eta_n]$ and the normal errors are scaled up (or down) proportionately to the scale of the $\beta$ vector.

The sixth and final model we consider is the "mixed-mixed-logit" or MM-MNL model. This generalizes N-MIXL by specifying $\beta_n$ in (1) as a discrete mixture-of-multivariate normals. The motivation for this model is that a mixture-of-normals provides a very flexible heterogeneity distribution. Ferguson (1973) shows the mixture-of-normals can approximate any heterogeneity distribution arbitrarily well, and Geweke and Keane (1999, 2001, 2007) show that a small number of normals can usually approximate even highly non-normal distributions quite well in practice. And MM-MNL has been shown to fit choice behavior better than N-MIXL in some recent studies (Rossi, Allenby and McCulloch (2005), Burda, Harding and Hausman (2008)).

Note that G-MNL and MM-MNL are related, as G-MNL assumes $\beta_n$ is a *continuous* mixture of scaled normals, while MM-MNL assumes it is a *discrete* mixture-of-normals. Also note that MM-MNL nests N-MIXL and LC, just as G-MNL nests N-MIXL and S-MNL.

The reason we consider this set of models is they are either very popular in applications (N-MIXL, LC) or have received considerable attention in recent work (MM-MNL, G-MNL, S-MNL). Keane and Wasi (2012a) compare performance of these models using data from ten stated preference (SP) choice experiments for a wide range of products. Based on the Bayes information criterion, they find MM-MNL is preferred in 4 datasets, while G-MNL and S-MNL are preferred in 3 each. Basically, MM-MNL is preferred in datasets with the most complex heterogeneity structures, while S-MNL is preferred in datasets with the simplest structures. They also found that the popular N-MIXL model was not preferred in any dataset,[3] and that LC performed poorly in all 10 datasets.[4]

However, in an application to supermarket scanner data on pizza, Keane and Wasi (2012b) found a very different ranking of the heterogeneity models. Specifically, MM-MNL was preferred over both G-MNL and N-MIXL, while the latter two models produced a similar fit. Both S-MNL and LC performed much worse. This difference arose because the structure of heterogeneity was quite different in the revealed preference (RP) vs. stated preference (SP) data.

In summary, we have described six of the most important discrete choice models in use today (and 5 of the main ways to model heterogeneity). It is important to understand how use of randomly reduced choice sets affects parameter estimates in all of these models.

---

[3] They also considered a version of N-MIXL where each attribute coefficient can be either normal or lognormal, and one searches to find the best fitting combination. They called this "T-MIXL." This model was preferred in 2 out of 10 datasets, and it performed fairly well in general. However, we omit it here in the interest of space. The effect of randomly reduced choice sets in this model is unlikely to differ from that in the closely related N-MIXL model.

[4] This is consistent with findings of Elrod and Keane (1995) that LC tends to understate the degree of heterogeneity.

## III. Previous Studies on Logit with Many Alternatives

McFadden (1978) showed that one can consistently estimate parameters of MNL using randomly selected subsets of the full choice set. To see why, we need some notation. Let $j_n$ denote the observed choice of person $n$. Let $C$ denote the full choice set which has $J$ elements, and let $D_n$ denote a subset of $C$. This subset is randomly constructed *except* that it must contain $j_n$. Let $\pi(D_n|j_n)$ denote the probability that subset $D_n$ is constructed from all the possible subsets of $C$ that contain $j_n$. Finally, let $P(j|\theta^*,C,x_n)$ be the probability that $j$ is chosen from $C$, where $\theta^*$ is the true parameter vector and $x_n$ is the matrix of attributes of the $J$ alternatives.

Note that $\pi(D_n|j_n)$ is a function chosen by the researcher. For example, if $J=100$, then one choice is the following rule: include $j_n$ in $D_n$ and then chose 19 addition alternatives by sampling without replacement from the remaining 99 elements of $C$. Then $D_n$ is a hypothetical choice set with 20 elements. Here $\pi(D_n|j_n)$ is a constant over all possible $D_n$, and $\pi(D_n|k)=\pi(D_n|j)$ for $k,j\in D_n$. Also note that, as the chosen alternative is always included in $D_n$, we have $\pi(D|j_n)=0$ if $j_n\notin D$.

Now consider the hypothetical log-likelihood we would construct if each consumer $n=1,\ldots,N$ chose from the hypothetical choice set $D_n$. Let $U_j(\theta,x_{nj})$ denote the "deterministic" part of utility in the logit model, which excludes the additive *iid* extreme value error terms. Then the likelihood is given by the simple MNL formula:

$$LL_N(\theta) = \frac{1}{N}\sum_{n=1}^{N}\ln\frac{\exp[U_{j_n}(\theta,x_{n,j_n})]}{\sum_{j\in D_n}\exp[U_j(\theta,x_{nj})]} \tag{4}$$

Here we have suppressed the $t$ subscripts to conserve on notation (i.e., we consider a single choice occasion per consumer). Taking the expectation of (4) over realizations from the data generating process (and viewing the $x_n$ as fixed) we obtain:

$$E[LL_N(\theta)] = \frac{1}{N}\sum_{n=1}^{N}\sum_{k\in C}\sum_{D\subset C}P(k|\theta^*,C,x_n)\pi(D|k)\ln\frac{\exp[U_k(\theta,x_{nk})]}{\sum_{j\in D}\exp[U_j(\theta,x_{nj})]} \tag{5}$$

It is important to note that $\pi(D|k)=0$ if $k\notin D$, and that the third summation is over all possible $D\subset C$, regardless of whether they have positive probability. Next, we multiply and divide $P(k|\theta^*,C,x_n) = \exp[U_k(\theta^*,x_{nk})]\Big/\sum_{j\in C}\exp[U_j(\theta^*,x_{nj})]$ by $\sum_{j\in D}\exp[U_j(\theta^*,x_{nj})]$. This gives:

$$E[LL_N(\theta)] = \frac{1}{N}\sum_{n=1}^{N}\sum_{k\in C}\sum_{D\subset C}\frac{\sum\limits_{j\in D}\exp[U_j(\theta^*,x_{nj})]}{\sum\limits_{j\in C}\exp[U_j(\theta^*,x_{nj})]}\frac{\exp[U_k(\theta^*,x_{nk})]}{\sum\limits_{j\in D}\exp[U_j(\theta^*,x_{nj})]}\pi(D\mid k)\ln\frac{\exp[U_k(\theta,x_{nk})]}{\sum\limits_{j\in D}\exp[U_j(\theta,x_{nj})]} \quad (6)$$

It is convenient to define $R_n(D,C,\theta^*) \equiv \sum\limits_{j\in D}\exp[U_j(\theta^*,x_{nj})]\Big/\sum\limits_{j\in C}\exp[U_j(\theta^*,x_{nj})]$.[5] Now, as

$R_n(D,C,\theta^*)$ does not depend on $k$ (but only on $D$) we can bring the sum over $k$ inside to obtain:[6]

$$E[LL_N(\theta)] = \frac{1}{N}\sum_{n=1}^{N}\sum_{D\subset C}R_n(D,C,\theta^*)\cdot\sum_{k\in C}\left\{\frac{\exp[U_k(\theta^*,x_{nk})]}{\sum\limits_{j\in D}\exp[U_j(\theta^*,x_{nj})]}\pi(D\mid k)\ln\frac{\exp[U_k(\theta,x_{nk})]}{\sum\limits_{j\in D}\exp[U_j(\theta,x_{nj})]}\right\} \quad (7)$$

Next we multiply and divide by $\pi(D|j)$, utilizing what McFadden (1978) calls the "uniform conditioning" property of $\pi$ – i.e., the fact that $\pi(D|k)=\pi(D|j)=\pi(D)$ if $k,j\in D$ – to obtain:

$$E[LL_N(\theta)] = \frac{1}{N}\sum_{n=1}^{N}\sum_{D\subset C}R_n(D,C,\theta^*)\pi(D)\cdot\sum_{k\in C}\left\{\frac{\exp[U_k(\theta^*,x_{nk})]\pi(D\mid k)}{\sum\limits_{j\in D}\exp[U_j(\theta^*,x_{nj})]\pi(D\mid j)}\ln\frac{\exp[U_k(\theta,x_{nk})]}{\sum\limits_{j\in D}\exp[U_j(\theta,x_{nj})]}\right\}$$

Now focus on the term $\exp[U_k(\theta^*,x_{nk})]\pi(D\mid k)\Big/\sum\limits_{j\in D}\exp[U_j(\theta^*,x_{nj})]\pi(D\mid j)$. Because $\pi(D|k)=0$ if

$k\notin D$, the numerator vanishes in all cases except where $k\in D$. Also, because $\pi(D|k)=\pi(D|j)$ if

$k,j\in D$, the $\pi$ terms cancel out and we are left with:

$$E[LL_N(\theta)] = \frac{1}{N}\sum_{n=1}^{N}\sum_{D\subset C}R_n(D,C,\theta^*)\pi(D)\cdot\sum_{k\in D}\left\{\frac{\exp[U_k(\theta^*,x_{nk})]}{\sum\limits_{j\in D}\exp[U_j(\theta^*,x_{nj})]}\ln\frac{\exp[U_k(\theta,x_{nk})]}{\sum\limits_{j\in D}\exp[U_j(\theta,x_{nj})]}\right\} \quad (8)$$

Notice that the third summation term has the form $\sum\limits_{k}\{P_k(\theta^*)\ln P_k(\theta)\}$. For any sequences of numbers

[5] Note that $R_n(D,C,\theta^*)$ is the ratio of the probability of choosing $k\in D$ from full choice set $C$ to that of choosing $k$ from the reduced choice set $D$ (That is, the ratio of the denominator of the logit choice probability for the reduced choice set $D$ to the full choice set $C$). Clearly $R_n(D,C,\theta^*)<1$ as $D\subset C$.

[6] At this point in the proof it is crucial to note that the third summation in (6) is over all possible $D\subset C$, regardless of whether they have positive probability. Specifically, the set of $D$ that is summed over here is not constrained by the fact that any $D$ that does not contain $k$ has zero probability. This is dealt with via the $\pi(D|k)$ term.

$\{P_1,....,P_J\}$ and $\{P_1^*,....,P_J^*\}$ such that $\sum_k P_k = 1$, $\sum_k P_k^* = 1$, and $P_k > 0$, $P_k^* > 0$ for all $k$, where the

$P_k^*$ are given and the $P_k$ are to be chosen, it is simple to show that $\sum_k \{P_k^* \ln P_k\}$ is maximized by

setting $P_k = P_k^*$ for all $k$.[7] In equation (8) the $P_k$ and $P_k^*$ correspond to the logit probability

expressions, and we achieve equality (for all $k$) by setting $\theta = \theta^*$. This completes the proof.

Returning to (4), note that the correct log-likelihood for the MNL model can be written:

$$LL_N^C(\theta) = \frac{1}{N}\sum_{n=1}^{N} \ln \frac{\exp[U_{j_n}(\theta, x_{n,j_n})]}{\sum_{j\in D_n}\exp[U_j(\theta, x_{nj})]} R_n(D,C,\theta) = \frac{1}{N}\sum_{n=1}^{N} \ln \frac{\exp[U_{j_n}(\theta, x_{n,j_n})]}{\sum_{j\in D_n}\exp[U_j(\theta, x_{nj})]} + \frac{1}{N}\sum_{n=1}^{N} \ln R_n(D,C,\theta)$$

And so the pseudo-log likelihood based on the reduced choice set is:

$$LL_N(\theta) = LL_N^C(\theta) - \frac{1}{N}\sum_{n=1}^{N} \ln R_n(D,C,\theta) = LL_N^C(\theta) + \frac{1}{N}\sum_{n=1}^{N} \ln\left[R_n(D,C,\theta)^{-1}\right] \qquad (9)$$

Thus, the basic intuition of McFadden's result is that the $(1/N) \sum \ln R_n(D,C,\theta)^{-1}$ term shifts the

(expected) log-likelihood up, but it does not alter where it is maximized. Notice that:

$$\frac{1}{N}\sum_{n=1}^{N} \ln R_n(D,C,\theta)^{-1} \equiv \frac{1}{N}\sum_{n=1}^{N} \ln \sum_{j\in C}\exp[U_j(\theta, x_{nj})] \Big/ \sum_{j\in D_n}\exp[U_j(\theta, x_{nj})] \qquad (10)$$

As the set $D_n$ contains the true choice (along with randomly selected other options from $C$), the

expectation of (10), i.e., the <u>positive</u> divergence between $LL_N(\theta)$ and $LL_N^C(\theta)$, is <u>minimized</u> at $\theta^*$.

Unfortunately, McFadden (1978)'s method of proof does not go through when MNL is

extended to include heterogeneity. To see why, consider a case with $T$ types of consumers with

type proportions $p^*(\tau)$ and parameters $\theta_\tau^*$ for $\tau = 1,...,T$. In that case equation (5) becomes:

$$E[LL_N(\theta)] = \frac{1}{N}\sum_{n=1}^{N}\sum_{k\in C}\sum_{D\subset C} P(k|\theta^*, p^*, C, x_n)\pi(D|k)\ln\left\{\sum_{\tau=1}^{T} p(\tau)\frac{\exp[U_k(\theta_\tau, x_{nk})]}{\sum_{j\in D}\exp[U_j(\theta_\tau, x_{nj})]}\right\} \qquad (11)$$

The term in brackets is the unconditional choice probability for person $n$, obtained by taking a

[7] For example, take $y = p^*\ln p + (1-p^*)\ln(1-p)$. Then the maximum is found by setting $dy/dp = p^*/p - (1-p^*)/(1-p) = 0$ which implies that $p = p^*$.

weighted sum of the choice probabilities conditional on type $\tau$, and weighting by the estimated type proportions, which we denote by $p(\tau)$. The term $P(k \mid \theta^*, p^*, C, x_n)$ is the true unconditional probability that option $k$ is chosen. This depends on the vector of true type proportions, denoted by $p^*$, and the vector of true parameters for each type $\theta^* = (\theta_1^*, \ldots, \theta_T^*)$. Specifically, we have:

$$P(k \mid \theta^*, p^*, C, x_n) = \sum_{\tau=1}^{T} p^*(\tau) \exp[U_k(\theta_\tau^*, x_{nk})] \Big/ \sum_{j \in C} \exp[U_j(\theta_\tau^*, x_{nj})]$$

Now consider the key step of the previous proof where we multiply and divide this object by $\sum_{j \in D} \exp[U_j(\theta_\tau^*, x_{nj})]$, where now the $\theta^*$ are type specific. This gives:

$$E[LL_N(\theta)] = \frac{1}{N} \sum_{n=1}^{N} \sum_{k \in C} \sum_{D \subset C} \sum_{\tau=1}^{T} R_n(D, C, \theta_\tau^*) p^*(\tau) \frac{\exp[U_k(\theta_\tau^*, x_{nk})]}{\sum_{j \in D} \exp[U_j(\theta_\tau^*, x_{nj})]} \pi(D \mid k) \ln\left\{ \sum_{\tau=1}^{T} p(\tau) \frac{\exp[U_k(\theta_\tau, x_{nk})]}{\sum_{j \in D} \exp[U_j(\theta_\tau, x_{nj})]} \right\}$$

where $R_n(D, C, \theta_\tau^*) \equiv \sum_{j \in D} \exp[U_j(\theta_\tau^*, x_{nj})] \Big/ \sum_{j \in C} \exp[U_j(\theta_\tau^*, x_{nj})]$ is the ratio of the probability of choosing $k \in D$ from the full choice set $C$ to that of choosing it from the reduced choice set $D$. In general this is type $\tau$ specific. However, in the special case $R_n(D, C, \theta_\tau^*) = R_n(D, C)$ for all $\tau$, we could use manipulations like those leading from (6) to (8) to obtain:

$$E[LL_N(\theta)] = \frac{1}{N} \sum_{n=1}^{N} \sum_{D \subset C} R_n(D, C) \pi(D) \cdot \sum_{k \in D} \left\{ \sum_{\tau=1}^{T} p^*(\tau) \frac{\exp[U_k(\theta_\tau^*, x_{nk})]}{\sum_{j \in D} \exp[U_j(\theta_\tau^*, x_{nj})]} \ln \sum_{\tau=1}^{T} p(\tau) \frac{\exp[U_k(\theta_\tau, x_{nk})]}{\sum_{j \in D} \exp[U_j(\theta_\tau, x_{nj})]} \right\}$$

In that case, the third summation term would have the form $\sum_k \{P_k(\theta^*, p^*) \ln P_k(\theta, p)\}$, and by the same logic as above the maximum is achieved by setting $\theta = \theta^*$ and $p = p^*$.

But, in general, when $R_n(D, C, \theta_\tau^*)$ is type specific, we cannot bring this term outside the summation over $\tau$, and so the form $\sum_k \{P_k(\theta^*, p^*) \ln P_k(\theta, p)\}$ cannot be achieved. Instead we obtain:

$$E[LL_N(\theta)] = \frac{1}{N} \sum_{n=1}^{N} \sum_{D \subset C} \pi(D) \sum_{k \in D} \left\{ \sum_{\tau=1}^{T} R_n(D, C, \theta_\tau^*) p^*(\tau) \frac{\exp[U_k(\theta_\tau^*, x_{nk})]}{\sum_{j \in D} \exp[U_j(\theta_\tau^*, x_{nj})]} \ln \sum_{\tau=1}^{T} p(\tau) \frac{\exp[U_k(\theta_\tau, x_{nk})]}{\sum_{j \in D} \exp[U_j(\theta_\tau, x_{nj})]} \right\}$$

So, unlike in (8), the necessary symmetry of the term in curly brackets is not achieved.

The basic problem here is that the subset $D$ is not chosen completely at random. It must include the chosen alternative $k$. Thus, $R_n(D,C,\theta_\tau^*)$ differs across types because $\exp[U_k(\theta_\tau^*, x_{nk})]$ differs across types. To give an extreme example, suppose that one type has an extreme preference for product type $k$, and purchases it with a probability close to one. Then for that type we have $R_n(D,C,\theta_\tau^*) \approx 1$. On the other hand, suppose a type makes choices randomly, choosing each product type with probability $1/(\#C)$. For that type we have $R_n(D,C,\theta_\tau^*) = (\#D)/(\#C)$.[8]

However, note that as the size of the randomly selected subset $D$ approaches that of the full choice set $C$, we have that $R_n(D,C,\theta_\tau^*) \to 1$ for all $\tau$. This suggests that if the random subsets $D$ are sufficiently large, then any bias induced by using random subsets of the full choice set $C$ will tend to be negligible. In the next section we examine this issue using a Monte Carlo study. We find that finite sample bias is indeed negligible even for modest sized subsets.

A few studies have examined empirically the impact of randomly reducing choice set size in MNL models with heterogeneity (MIXL models). For instance, Brownstone, Bunch and Train (2000) estimate an N-MIXL model of auto demand (with 5 random coefficients). Their dataset is cross-sectional and combines RP and SP data (with 607 and 4,656 observations, respectively). The RP choice set contains 689 makes/models of cars (and more than 20 attributes). They apply a sampling procedure that reduces the choice set to only 28 (4%). They report that results do not change systematically when they increase the size of the sampled choice set.[9]

McConnell and Tseng (2000) examine a relatively small choice problem for 388 beach users with one trip each. The beaches have 4 attributes, one of which has a random coefficient. The authors compare estimates obtained using the full choice set of 10 choices vs. those obtained using a sampled subset of only 4 choices, and find that the results do not significantly change.

Narella and Bhat (2004) perform a Monte Carlo study on an N-MIXL model with a choice set of 200 alternatives, using different sizes for the sampled subsets. Their use a simulated cross-section of 750 people. Choices have 5 attributes of which 2 have random coefficients. They suggest that one can obtain reliable estimates using a 25% sub-sample of the full choice set.

Domanski and von Haefen (2010), in contrast, consider the latent class (LC) model. In their application the full choice set consists of 569 lakes in Wisconsin. The data is an unbalanced

---

[8] If the sampling scheme could be designed so $R_n(D,C,\theta_\tau) = (\#D)/(\#C)$ for all types, then we would have $R_n(D,C,\theta_\tau) = R_n(D,C)$ and McFadden's proof goes through. Narella and Bhat (2004) also discuss this (see their equation (8)).
[9] However, this stability with respect to the size of the RP data choice set may arise simply because the likelihood contribution of the RP data is much smaller than that of the SP data in their application.

panel of 513 respondents whose number of trips range from 1 to 50. There are 15 attributes. They estimate the LC model via the EM algorithm using both the full choice set and randomly sampled subsets ranging from 50% to 1% of the full set. They find one can obtain reasonably reliable estimates using 5% subsets.[10]

These results suggest that the bias induced by using randomly selected subsets of the full choice set may be minor in MNL models, even with heterogeneity. However, it is difficult to generalize from these authors' results to the types of models and data that we emphasized in Section II. For instance, the earlier papers that use the N-MIXL model only consider cross-section data. But our interest is primarily in the panel data case.[11] Only Domanski and von Haefen (2010) consider panel data, but they only consider the LC model, not the other models that we discussed in Section II (N-MIXL, MM-MNL, G-MNL and S-MNL). Therefore, based on prior studies, it is hard to assess the bias from using random subsets of the full choice set for the types of models and data we are interested in.

Thus, in the next section, we turn to our own Monte Carlo study of N-MIXL, MM-MNL, and G-MNL models in the panel case. [We do not consider S-MNL separately as it is a special case of G–MNL]. Then, in Section V, we consider a real data application where we look at supermarket scanner data on demand for frozen pizza. In this application the complete choice set contains nearly 100 alternatives, and we compare results for different sized subsets of the full set.

## IV. Monte Carlo Simulation

Here, we report results of Monte Carlo experiments to assess the bias induced by using randomly selected subsets of the full choice set to estimate N-MIXL, G-MNL and MM-MNL models. In each experiment, there are 200 hypothetical respondents and 20 choice occasions per respondent (4,000 observations). There are 60 alternatives in the full choice set, each with 4 attributes. The first and second attributes are dummy variables.[12] The third and fourth are drawn from standard normal distributions. The experiments differ in the specification of heterogeneity. We generate 20 artificial datasets for each case.

We estimate each model using both the full choice set (60 alternatives) and randomly sampled subsets with 20 or 10 alternatives. The chosen alternative is always included. Then,

---

[10] Also, von Haefen and Jacobsen (unpublished) present a series of Monte Carlo experiments showing that the finite sample bias from using randomly selected subsets of the full choice set in LC models is typically small.

[11] This is because, in our view, the estimation of individual level parameters is most relevant in the panel data case.

[12] Each of these dummies is equal to 1 with probability 0.5.

either 19 or 9 additional alternatives are randomly drawn from the remaining 59 alternatives. The random choice sets are drawn independently for each observation (person). Thus, people with the same observed choices will have random choice sets with different sampled alternatives. In the estimation, 500 draws are used to simulate the likelihood.

First, consider N-MIXL applied to artificial datasets of different design. Table 1 reports a case where only the first of the four attributes has a random coefficient, with a mean and standard deviation of one. The table reports the true parameter values, the mean estimates across the 20 Monte Carlo data sets, the empirical standard deviation of the estimates, and the mean of the asymptotic standard errors using both the BHHH (outer product) and robust (sandwich) formulas. An asterisk indicates <u>bias</u> in an estimated parameter is significant at the 5% level.

In Table 1 there is no evidence of significant bias using the full choice set of 60 or a random subset of 20 alternatives. If we use a random subset of only 10 alternatives there is significant bias only for $\beta_3$. But the bias is quantitatively small, as the true value is 1.0 and the mean estimate is 1.01. Also, the estimated and empirical standard errors align closely. Table 2 reports results for a case with (correlated) random coefficients on the first two attributes. The results are almost identical to those in Table 1.

Table 3 considers N-MIXL with a full variance-covariance matrix. Here, only a few covariance matrix parameters exhibit significant bias when the choice set is reduced from the full set of 60 to either 20 or 10. Again the magnitude of the bias is quantitatively small. Here, however, the BHHH standard errors are small relative to the empirical standard errors, regardless of whether the full or reduced choice set is used. Robust standard errors are more accurate.

We consider the G-MNL model with a full covariance matrix in Table 4. Here there is little evidence of bias, regardless of whether we use the full choice set or subsets of 20 or 10. In each case, we see significant (but quantitatively small) bias for just one covariance parameter. This is not surprising given this model has 16 parameters. Again however, BHHH standard errors using only 500 draws are too small, and robust standard errors are more accurate.

Finally, Table 5 reports results for the MM-MNL model. We consider a case where there are two consumer types, each with its own mean vector and covariance matrix for the parameter vector (the covariance matrices are assumed diagonal). There is no evidence of bias when using the full choice set. If we use a subset of 20 choices there is a significant but quantitatively small bias for the standard deviation of the $\beta_2$ parameter for type 1 (0.94 vs. 1.0). And if we use a

12

subset of 10 choices there is a significant but small bias for the standard deviation of the $\beta_2$ parameter for type 2 (0.51 vs. 0.60). The BHHH standard errors are again too small. However, we re-calculated the BHHH standard errors in Tables 3-5 using 5000 draws, and the results were much more accurate. Thus, in the more complicated cases of Tables 3-5, we find that 500 draws is too few to give reliable estimates of the standard errors. But if we use 5000 draws the empirical and asymptotic standard errors align well.

In summary, we find little evidence that the use of randomly selected subsets of the full choice set induces bias in estimates of the N-MIXL, G-MNL or MM-MNL models. This is despite the fact that the proof of consistency of logit models estimated on reduced choice sets does not go through when there is heterogeneity (see Section III). In the next subsection we give some intuition for this finding.

### IV.A. Some Intuition for the Monte Carlo Results

As we discussed in Section III, the McFadden (1978) proof would go through even in the case of heterogeneity if the condition $R_n(D,C,\theta_\tau^*) = R_n(D,C)$ for all types $\tau$ were to hold. As we noted, this condition would hold if the subset $D$ were chosen completely at random, but that is not the case, as $D$ must include the chosen alternative $k$. Nevertheless, we conjecture that, while this condition does not hold exactly, it holds to a very good approximation. And as a result, the bias from using reduced choice sets is negligible.

The accuracy of the approximation is illustrated in Figure 1. It is based on the experiment in Table 1 that considered an N-MIXL model where only attribute #1 has a random coefficient. The dashed line is the log-likelihood contour as we vary $\hat{\beta}_1$ holding the other parameters at the true values. The solid line is the pseudo-log-likelihood contour if we use a random subset of 10 alternatives instead of the full choice set. (The log-likelihood value at each $\hat{\beta}_1$ is an average from 20 simulated datasets). Note that these two likelihood contours are plotted on different scales (see the left and right axes) so they can fit in the same graph. The bottom panel of Figure 1 plots the (positive) difference between the log-likelihoods based on the reduced vs. full choice sets.

As we noted earlier, the basic intuition of McFadden's consistency result is that the use of a random subset of the full choice set shifts the (expected) log-likelihood up, but does not alter where it is maximized. Even though this result does not hold exactly with heterogeneity, it holds to a very good approximation, as can be seen clearly in Figure 1. The true and pseudo-likelihood contours have very similar shapes and are maximized at almost the same point (1.02 vs. 1.01).

The difference between the true and pseudo-likelihoods (bottom panel) varies by only about ½ point over a very wide range around the optimum. Hence it has little effect on the optimum.

## IV.B. The Bias in Information Criteria

An important point is that, even if using reduced choice sets does not induce significant bias in the estimates, it will lead to bias in information criteria used to compare models. Consider the Bayes Information Criterion, BIC = $-2LL + k \cdot \ln(N)$. The second term is the penalty for parameters ($k$), which depends on the number of observations N. If we (randomly) reduce the size of the choice set, the penalty term remains the same, but the (pseudo)-log-likelihood will improve. So the BIC difference between two models is only invariant to the size of the random choice set under the stringent condition that the log-likelihood difference between the two models is the same for all random choice set sizes. But in our simulations this is not the case.

To illustrate this point, we take the simulated data sets used to study the N-MIXL model in Table 2 and estimate simple MNL models on all of them. In Table 6, top panel, first two columns, we report the average log-likelihood for MNL (the misspecified model) and N-MIXL (the true model). The next column reports the difference between the log-likelihoods. Note that the difference is <u>smaller</u> for smaller random choice set sizes. As a result, the smaller the random choice set, the smaller the BIC gain from estimating the true model. We see the same pattern in the bottom panel, which compares MNL and MM-MNL (using data sets from Table 5). Thus, we see how BIC is biased toward smaller models when only a subset of alternatives is used in the estimation. This should be kept in mind when comparing models in the next section.

## V. Estimation Results for Scanner Data on Frozen Pizza

In this section, we present an empirical application to demand for frozen pizza, where there are nearly 100 varieties to choose from. Before turning to the substantive results, we first study the effect of using random subsets of the full choice set in this context.

We use data on household purchases of frozen pizza collected by IRI at a large store in Eau Claire, Wisconsin, from Jan. 2001-Dec. 2003. In these data, transactions are observed at the Universal Product Code (UPC) level. Over 400 UPCs of frozen pizza are recorded during the three years. But this is only because UPC codes change due to very minor changes in ingredients, packaging or weight. Thus, instead of defining choices at UPC level, we choose to define "types" of pizza. Types are characterized by more significant product characteristics such as brand, topping, type of crust, and whether the pizza can be heated by microwave.

There are 102 types of pizza available sometime during the study period. But within any one week the number of available types varies from 72 to 96. These large choice set sizes make computation burdensome for discrete choice models with heterogeneity, particularly given that we have a large scanner dataset with many time periods.

We apply two sample screens: First, only panelists who IRI classifies as consistent reporters are included. We also require that sample members be regular frozen pizza consumers. Specifically, they must make 15 to 60 purchases in the 3 years. There are 129 panelists who meet our criteria. The total number of shopping trips where frozen pizza is bought is 4,123, an average of 32 purchases per person.[13] (Note: People buy frozen pizza on roughly 30% of all shopping trips. We do not attempt to model category purchase timing decisions, as that is beyond the scope of the present paper; see Ching et al (2009) for further discussion).

In each model, the characteristics we use to predict utility from each type of pizza are brand (5 major brands plus "other"), topping (7 types), crust (3 types), if the pizza is micro-waveable, price and if the pizza is on promotion (i.e., on feature or display). The construction of the price and promotion variables deserves additional comment. For the pizza a consumer actually buys, the recorded price and promotion information for that UPC is used. For the non-chosen pizza types, the price and promotion variables are a weighted average over all UPCs within that <u>type</u>. The weights are the market shares for the whole period the IRI data is available.

Two of the models we consider have closed form choice probabilities (MNL and LC) while four do not because they have continuous heterogeneity distributions (N-MIXL, G-MNL, S-MNL and MM-MNL). We estimate the latter four models using simulated maximum likelihood (SML), based on 500 draws to simulate the likelihood. As we found in Section IV, this is not sufficient to provide reliable estimates of the standard errors. Thus, we use 500 draws in the estimation process, but then shift to using 5000 draws when we calculate the standard errors at the final estimates.

Several of our models have alternative versions. For example, in LC and MM-MNL one has to choose the number of latent types. In G-MNL, N-MIXL and G-MNL one has to decide if errors are correlated. In each case, we try several versions and report the one preferred by BIC.

---

[13] People buy only one type of pizza on 50 percent of purchase occasions. In cases where consumers buy $K$ types, we treat each purchase as a separate observation. But we take the geometric mean of the likelihood contributions. $(\prod_k p_k)^{1/K}$. For MNL this is the same as weighting the log-likelihood contributions by $1/K$. Less than 2 percent of trips involve purchase of more than 4 types. For these trips, we randomly select 4 types to include in the analysis.

## V.A. Comparing Results Using Different Random Choice Set Sizes

In Tables 7 to 9 we report estimates obtained using 3 different randomly reduced choice set sizes, of 20, 30 or 40, respectively. A striking finding is that estimates are quite stable across choice set sizes for all 6 models (MNL, S-MNL, N-MIXL, G-MNL, LC, and MM-MNL). This is obviously not surprising for MNL, which is consistent for randomly chosen subsets of the full choice set. But it is surprising for models with heterogeneity.

For example, for N-MIXL, the price coefficients are -1.22 (standard error = 0.07), -1.18 (0.06) and -1.21 (0.06) using choice sets of 20, 30 and 40, respectively. So, variation of the estimates across the three choice set sizes is less than one standard error. Similarly, in G-MNL, the price coefficients are -1.23 (0.07), -1.16 (0.06) and -1.20 (0.06), respectively. The MM-MNL model identifies two types in each case. Price coefficients for type 1 are -1.62 (0.11), -1.66 (0.11) and -1.75 (0.07), while those for type 2 are -0.26 (0.14) -0.33 (0.11) and -0.48 (0.09). Results for LC and S-MNL are similar. In all cases, the price coefficient is stable across choice set sizes.

The reader can verify that the other parameters besides price are also quite stable across the three choice set sizes.[14] This similarity adds to our confidence (already considerable in light of the earlier Monte Carlo results) that any bias induced by using random subsets of the full choice set is negligible, even for models that include heterogeneity.[15]

Finally, note that BIC prefers the mixture-of-normals model (MM-MNL) by 230 to 300 points over G-MNL (with N-MIXL a close third). This is consistent with results in Rossi et al (2005) and Burda et al (2008). Table 10 compares predicted market shares for brand/topping combinations from the MM-MNL model with the actual shares. Clearly, the fit is rather good.

## V.B. Substantive Results from the Estimation

It has been common in marketing to use scanner data to estimate price elasticities of demand at the brand level. If retailers move the prices of all varieties of a brand in tandem, then this is all that is of interest. But inspection of our data reveals the existence of a great deal of price variability for individual varieties *within* brands. Our use of a large choice set, with varieties treated as separate options, enables us to compare brand vs. variety level elasticities, and to decompose variety level elasticities into brand vs. variety components. (To keep things manageable, we will aggregate varieties to the brand/topping level, where there are 37 options).

---

[14] The identity of segments one and two flips between the LC models with 20 or 40 choices and that with 30 choices. But this is not too surprising given that the two largest segments are very close in size.
[15] We even tried random choices sets of only 5. The estimates continued to look reasonable, although differences from the models with larger choice sets became somewhat noticeable. This was true even for MNL.

First, Table 11 reports brand level price elasticities for selected brands. In forming these elasticities, we assume prices of all varieties of a brand are increased proportionally. We report elasticities for all 6 models, estimated using random choice set sizes of 20, 30 and 40. The first thing to notice is that elasticities are quite stable across the different random choice set sizes. This is consistent with our previous finding that the parameter estimates are quite stable.

For Tombstone, which is the largest brand (26% market share), the own price elasticity is -1.66 according the preferred MM-MNL model (estimated using random choice sets of size 40). The simple MNL model implies a nearly identical elasticity of -1.71. So, for this purpose, it is not clear why one would go to the trouble of estimating the MM-MNL model.

But the story is different if we look at Jacks, which has a smaller market share (20%). Here the MM-MNL model gives an elasticity of -2.14, while MNL gives -1.69. Thus, MM-MNL implies a larger price elasticity of demand for the smaller brand, and MNL implies the reverse. Indeed, this is a fairly general pattern. Across all brands/varieties, the correlation between the log market share and the price elasticity is .53 for MM-MNL (i.e., higher share, smaller elasticity), while the correlation implied by MNL is -.50 (i.e., higher share, larger elasticity). This is a fundamental behavioral difference between the models. Furthermore, all the models with heterogeneity generate a positive correlation (i.e., higher share, smaller elasticity).

In Table 12 we report variety-specific elasticities for brand/topping combinations. We only report results for the preferred MM-MNL model (estimated using a random choice set size of 40). First, in the top panel we consider Tombstone with the sausage/pepperoni topping. This is the most popular of all varieties, with a market share of 11%. The price elasticity of demand for this variety is -1.68, compared to -1.66 for the whole Tombstone brand.

The closeness of the brand and variety level elasticities at first seems very surprising. Conventional wisdom suggests that the price elasticity for one variety should be appreciably larger than that for a whole brand, because a variety has more close substitutes. That is, in response to a price increase for a single variety, a consumer who is loyal to the Tombstone brand can switch to other varieties of Tombstone, rather than having to switch a different brand.

What drives the result becomes apparent if we decompose the brand level elasticity at the variety level. When the whole Tombstone brand raises its price, the elasticity for the sausage variety is -1.35, while that for all other varieties combined is -1.93. Thus, the (percentage) impact of the price increase on the less popular varieties is substantially greater than the impact on the

17

most popular variety. So the price elasticity for the sausage variety is indeed quite a bit larger with respect to its own price than with respect to the price of the whole brand (-1.68 vs. -1.35).

The situation is different when we look at varieties with a smaller market share. The average market share in the data is 100/37 = 2.7%. So consider Jacks meat/supreme pizza, which has a "typical" market share of 2.6%. In this case the variety level elasticity is -3.34, which is much greater than the elasticity of -2.14 for the Jacks brand as a whole. The compositional effect we saw with Tombstone does not arise here, because the elasticity of the Jacks meat variety with respect to the Jacks brand price is almost identical to that of all the other varieties.

Next we decompose variety level elasticities into their source components. We simulate that a 10% increase in the price of Tombstone sausage/pepperoni would cause its sales to drop by 17%. We can decompose this drop into three components: 27% is due to consumers switching to other varieties of Tombstone, 41% is due to switching to other brands while stilling choosing the sausage/pepperoni topping, and 32% is due to switching to other varieties of other brands.

For Jacks meat/supreme pizza we simulate that 10% price increase would cause sales to drop by 33%. Of this drop, 43% is due to consumers switching to other varieties of Jacks. Only 14% is due to consumers switching to other brands while still choosing the meat/supreme topping, and 43% is due to switching to other varieties of other brands.

Notice that in the case of the less popular topping (meat/supreme) we see much more switching to other toppings within the same brand (43% vs. 27%), and much less switching to other brands while staying with the same topping (14% vs. 41%). Both patterns seem intuitive. We also see a higher level of switching to completely difference brands/varieties (43% vs. 32%).

Viewed another way, we can decompose the variety level elasticities into the parts due to switching varieties (within brand), switching brand (within variety) and switching brand and variety. The elasticity of -1.68 for Tombstone sausage/pepperoni decomposes into 0.45, 0.69 and 0.54, respectively. The elasticity of -3.34 for Jacks meat/supreme decomposes into 1.43, 0.46 and 1.45, respectively. Clearly, the larger elasticity for the smaller market share variety is due to both (i) greater propensity to switch to other varieties within the same brand, (ii) greater propensity to switch to other brands/varieties.

For both the large and medium market share varieties, the preferred MM-MNL model implies that, after a price increase, the majority of switchers go to a different brand, rather than a different variety of the same brand (the fraction who completely switch brand are 73% and 57%

in the two cases). This is also true for small market share brands and small varieties.[16] Thus, our results suggest brand loyalty is not strong enough to keep the majority of switching consumers within the same brand if one variety of that brand increases its price. This finding highlights the importance of considering not only brands but also varieties when studying consumer choice behavior. This, in turn, highlights the importance of methods to deal with very large choice sets.

**VI. Conclusion**

Very large choice sets are a common problem in empirical work. A good example is the study of consumer demand in markets where there are many varieties of a product (as is usually the case at the UPC or SKU level). The use of random subsets of the full choice set in estimation is a potential solution to the problem of very large choice sets, but there is little evidence on whether such an approach is reliable in models with heterogeneous agents.

Here, we find evidence that using randomly reduced choice sets does not cause significant bias in estimating logit choice models with heterogeneity on very large choice sets. We presented both Monte Carlo evidence and results from supermarket scanner data to support this claim. Also, based on McFadden's (1978) original proof for the MNL case, we have given some intuition why his result is likely to hold approximately even if heterogeneity is introduced.

We also present a substantive application to choice among types of frozen pizza. A common procedure in marketing has been to aggregate varieties within brands in order to obtain pure brand choice problems where choice sets are much smaller. But we find that consideration of the full choice set, including the many varieties that exist within brands, can be very informative. For instance, our results suggest that brand loyalty is not strong enough to keep the majority of switching consumers within the same brand if one variety of that brand increases its price. This implies that preferences for varieties are also quite important.

Our finding that the use of random subsets of the full choice set is a reliable procedure is not only relevant for choice of supermarket goods. There are many contexts where large choice sets are a problem, such as choice of homes or residential location, choice of colleges and majors, choice of TV shows or movies, choice of occupation at a detailed level, etc.. So the use of randomly reduced choice sets may be useful in many contexts.

---

[16] For instance, we also looked at Red Baron bacon/burger, which has a share of only 0.47%. We simulate that 10% price increase would cause sales to drop by 29%. Of this drop, 24% is due to consumers switching to other varieties of Red Baron. Only 7% is due to consumers switching to other brands while still choosing the bacon/burger topping, and 69% is due to switching to other varieties of other brands. The total fraction who switch to other brands is 76%.

# References

Andrews, R.L. and A.K. Manrai (1999), "MDS Maps for Product Attributes and Market Response: An Application to Scanner Panel Data," *Marketing Science,* 18(4), 584-604.

Brownstone, D., D. Bunch and K. Train (2000), "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles," *Transportation Research B,* 34(5), 315-338.

Burda, M., M. Harding and J. Hausman (2008), A Bayesian mixed logit-probit model for multinomial choice. *Journal of Econometrics* 147: 232-246.

Ching, A., T. Erdem and M. Keane (2009), "The Price Consideration Model of Brand Choice," *Journal of Applied Econometrics*, 24:3 (March/April), 393-420.

Chintagunta, P. K. and J Dubé, (2005), "Estimating a Stockkeeping-Unit-Level Brand Choice Model That Combines Household Panel Data and Store Data," *Journal of Marketing Research*, 42(3), 368-379.

Domanski, A. and R.H. von Haefen (2010), "Estimation and Welfare Analysis from Mixed Logit Recreation Demand Models with Large Choice Sets," Working paper.

Dubé, J. (2004), "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science,* 23(1), 66-81.

Elrod, Terry and M.P. Keane (1995), "A Factor Analytic Probit Model for Representing the Market Structure in Panel Data," *Journal of Marketing Research*, 32, 1-16.

Fader, P.S. and B.G.S. Hardie (1996), "Modeling Consumer Choice among SKUs," *Journal of Marketing Research*, 33(4), 442-452.

Ferguson, T.S. (1973), A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1: 209-230.

Fiebig, D., M. Keane, J. Louviere and N. Wasi (2010), The Generalized Multinomial Logit Model: Accounting for scale and coefficient heterogeneity. *Marketing Science* 29, 393-421.

Geweke, J. and M. Keane (1999), Mixture of Normals Probit Models. in *Analysis of Panels and Limited Dependent Variable Models*, Hsiao, Lahiri, Lee and Pesaran (eds.), Cambridge University Press, 49-78.

Geweke, J. and M. Keane (2001), Computationally Intensive Methods for Integration in Econometrics. In *Handbook of Econometrics: Vol. 5*, J.J. Heckman and E.E. Leamer (eds.), Elsevier Science B.V., 3463-3568.

Geweke, J. and M. Keane (2007), Smoothly Mixing Regressions. *Journal of Econometrics* 138: 291-311.

Ho, T. and J. Chong (2003), "A Parsimonious Model of Stockkeeping-Unit Choice," *Journal of Marketing Research*, 40(3), 351-365.

Inman, J.J. , J. Park and A. Sinha (2008), "A Dynamic Choice Map Approach to Modeling Attribute-Level Varied Behavior Among Stockkeeping Units," *Journal of Marketing Research*,  45(1), 94-103.

Keane, Michael P. (1994), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95-116.

Keane, M.P. and N. Wasi (2012a), "Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior," *Journal of Applied Econometrics*, forthcoming.

Keane, M.P. and N. Wasi (2012b), "The Structure of Consumer Taste Heterogeneity in Revealed vs. Stated Preference Data," Working paper, University of New South Wales.

Louviere, Jordan J., Robert J. Meyer, David S. Bunch, Richard Carson, Benedict Dellaert, W. Michael Hanemann, David Hensher and Julie Irwin (1999), "Combining Sources of Preference Data for Modelling Complex Decision Processes," *Marketing Letters*, 10:3, 205-217.

Louviere, J.J., R.T. Carson, A. Ainslie, T. A. Cameron, J. R. DeShazo, D. Hensher, R. Kohn, T. Marley and D.J. Street (2002), "Dissecting the random component of utility," *Marketing Letters*, 13, 177-193.

McConnell, K.E. and W.Tseng (2000), "Some Preliminary Evidence on Sampling of Alternatives with the Random Parameters Logit," *Marine Resource Economic*s, 14, 317–332.

McCulloch, Robert and Rossi, Peter E. (1994), "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics*, Elsevier, 64(1-2), 207-240.

McFadden, D. (1974), Conditional Logit Analysis of Qualitative Choice Behavior, in *Frontiers in Econometrics*, in P. Zarembka (ed.), New York: Academic Press, 105-42.

McFadden, D. (1978), Modeling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam, 75–96.

McFadden, D. (1989), "A Method of Simulated Moments for the Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57:5, 995-1026.

McFadden, D. and K. Train (2000), "Mixed MNL models for discrete response," *Journal of Applied Econometrics,* 15, 447-470.

Meyer, Robert. J. and Jordan J. Louviere (2007), "Formal Choice Models of Informal Choices: What Choice Modelling Research Can (and Can't) Learn from Behavioral Theory", *Review of Marketing Research*, 4, (in press).

Narella, S. and C. Bhat (2004), "A Numerical Analysis of the Effect of Sampling of Alternatives in Discrete Choice Models," Transportation Research Record, 1894, pp. 11-19.

Pakes, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks, *Econometrica*, 54: 4 (July), pp. 755-784.

Rossi, P., Allenby, G. and R. McCulloch (2005), *Bayesian Statistics and Marketing*, John Wiley and Sons, Hoboken, N.J.

Singh, V.P., K. T. Hansen, and S. Gupta (2005), "Modeling Preferences for Common Attributes in Multicategory Brand Choice," *Journal of Marketing Research*, 42,195–209.

von Haefen, R. and M. Jacobsen (unpublished). "Sampling of Alternatives in a Mixture Model." Working Paper, Department of Agricultural and Resource Economics, North Carolina State University.

**Figure 1: Simulated MIXL Likelihood Profiles**
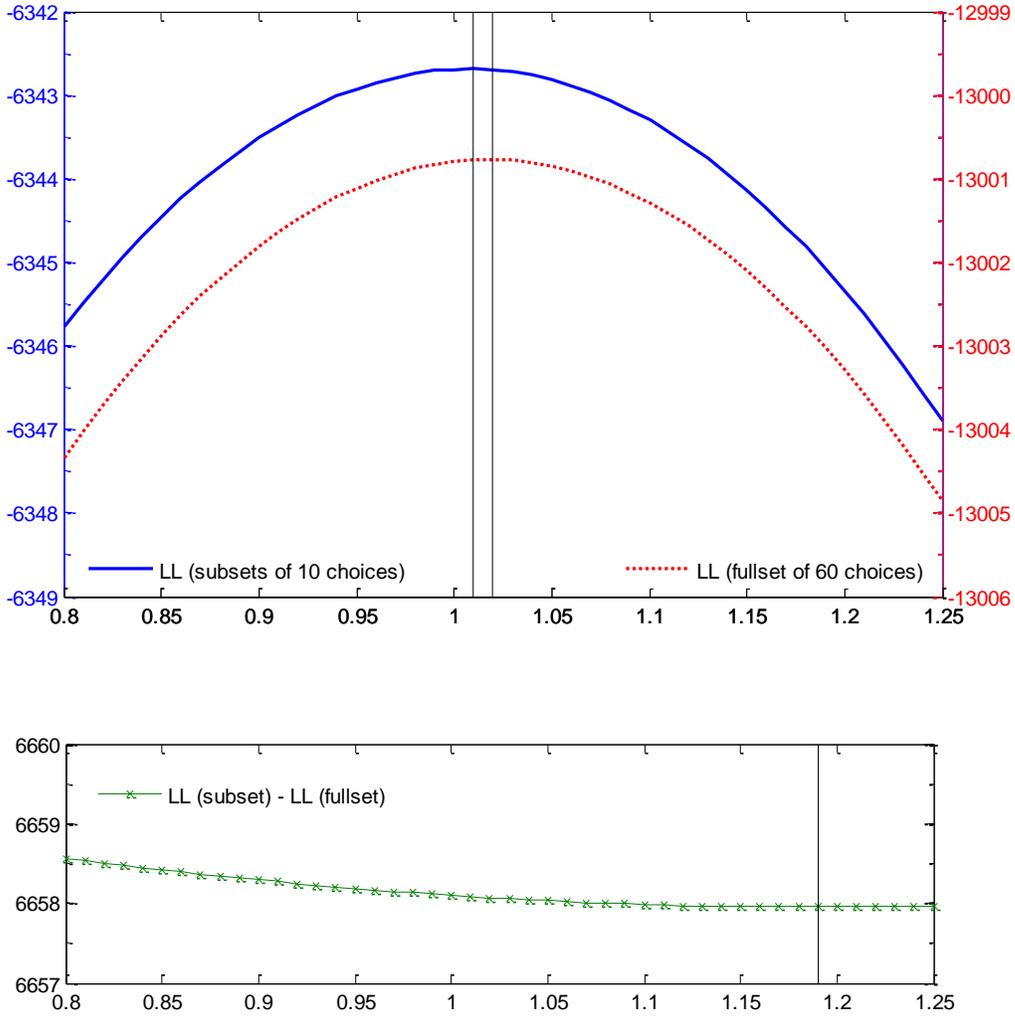


Note: In the top panel the red line is the log-likelihood profile for $\hat{\beta}_1$ based on the full choice set of J = 60. The blue line is the log-likelihood profile based on a random subset of 10 choices (including the chosen alternative). The scale for the blue line is on the left axis while that for the red line is on the right axis. The vertical lines indicate the maximum for each profile. The green line in the bottom panel is the (positive) difference between the two profiles.

**Table 1: Monte Carlo Results for Mixed Logit with One Random Coefficient**

| | | Complete choiceset (60 choices) | | | | Random subset (20 choices) | | | | Random subset (10 choices) | | | |
| | | mean | empi. | mean ASE | | mean | empi. | mean ASE | | mean | empi. | mean ASE | |
| | TRUE | est | s.e. | BHHH | Robust | est | s.e. | BHHH | Robust | est | s.e. | BHHH | Robust |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_1$ | 1 | **1.02** | 0.09 | 0.08 | 0.08 | **1.01** | 0.09 | 0.08 | 0.08 | **1.01** | 0.09 | 0.09 | 0.08 |
| $\beta_2$ | 1 | **1.00** | 0.04 | 0.04 | 0.04 | **1.00** | 0.04 | 0.04 | 0.04 | **1.01** | 0.05 | 0.04 | 0.04 |
| $\beta_3$ | 1 | **1.01** | 0.02 | 0.02 | 0.02 | **1.01** | 0.02 | 0.02 | 0.02 | **1.01** | 0.02* | 0.03 | 0.03 |
| $\beta_4$ | -1 | **-1.00** | 0.03 | 0.02 | 0.02 | **-1.01** | 0.03 | 0.02 | 0.02 | **-1.01** | 0.03 | 0.03 | 0.03 |
| $\sqrt{\sigma_{11}}$ | 1 | **1.01** | 0.10 | 0.07 | 0.07 | **1.00** | 0.10 | 0.08 | 0.08 | **0.99** | 0.09 | 0.08 | 0.08 |

Note: $\sqrt{\sigma_{11}}$ denotes standard deviation of $\beta_1$.

**Table 2: Monte Carlo Results for Mixed Logit with Two Correlated Random Coefficients**

| | | Complete choiceset (60 choices) | | | | Random subset (20 choices) | | | | Random subset (10 choices) | | | |
| | | mean | empi. | mean ASE | | mean | empi. | mean ASE | | mean | empi. | mean ASE | |
| | TRUE | est | s.e. | BHHH | Robust | est | s.e. | BHHH | Robust | est | s.e. | BHHH | Robust |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_1$ | 1 | **1.02** | 0.10 | 0.08 | 0.08 | **1.01** | 0.10 | 0.08 | 0.09 | **1.00** | 0.09 | 0.08 | 0.09 |
| $\beta_2$ | 1 | **1.02** | 0.07 | 0.08 | 0.09 | **1.02** | 0.07 | 0.08 | 0.09 | **1.01** | 0.07 | 0.09 | 0.09 |
| $\beta_3$ | 1 | **1.00** | 0.03 | 0.02 | 0.02 | **1.01** | 0.03 | 0.03 | 0.02 | **1.02** | 0.03* | 0.03 | 0.03 |
| $\beta_4$ | -1 | **-1.00** | 0.03 | 0.02 | 0.02 | **-1.01** | 0.03 | 0.02 | 0.03 | **-1.01** | 0.03 | 0.03 | 0.03 |
| $\sigma_{11}$ | 1 | **1.00** | 0.17 | 0.15 | 0.15 | **0.99** | 0.19 | 0.15 | 0.16 | **0.99** | 0.17 | 0.16 | 0.16 |
| $\sigma_{22}$ | 1 | **1.02** | 0.13 | 0.16 | 0.15 | **1.00** | 0.12 | 0.16 | 0.15 | **1.00** | 0.12 | 0.16 | 0.16 |
| $\sigma_{21}$ | 0.60 | **0.60** | 0.10 | 0.11 | 0.11 | **0.59** | 0.10 | 0.11 | 0.11 | **0.58** | 0.09 | 0.11 | 0.11 |

Note: $\sigma_{ii}$ denotes variance of $\beta_i$ and $\sigma_{ij}$ denotes covariance of $\beta_i$ and $\beta_j$.

**Table 3: Monte Carlo Results for Mixed Logit with Four Correlated Random Coefficients**

| | | Complete choiceset (60 choices) | | | | Random subset (20 choices) | | | | Random subset (10 choices) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | mean est | empi. s.e. | mean ASE | | mean est | empi. s.e. | mean ASE | | mean est | empi. s.e. | mean ASE | |
| | TRUE | | | BHHH | Robust | | | BHHH | Robust | | | BHHH | Robust |
| $\beta_1$ | 1 | 1.00 | 0.12 | 0.07 | 0.10 | 1.01 | 0.11 | 0.08 | 0.09 | 1.02 | 0.12 | 0.08 | 0.11 |
| $\beta_2$ | 1 | 1.01 | 0.10 | 0.07 | 0.11 | 1.00 | 0.12 | 0.08 | 0.10 | 1.01 | 0.11 | 0.08 | 0.10 |
| $\beta_3$ | 1 | 1.00 | 0.11 | 0.06 | 0.10 | 1.01 | 0.12 | 0.07 | 0.09 | 1.01 | 0.10 | 0.07 | 0.09 |
| $\beta_4$ | -1 | -1.02 | 0.12 | 0.06 | 0.10 | -1.01 | 0.10 | 0.07 | 0.09 | -1.01 | 0.11 | 0.07 | 0.10 |
| $\sigma_{11}$ | 1 | 1.00 | 0.23 | 0.14 | 0.16 | 1.01 | 0.22 | 0.15 | 0.17 | 0.98 | 0.21 | 0.16 | 0.17 |
| $\sigma_{22}$ | 1 | 1.04 | 0.18 | 0.15 | 0.17 | 1.04 | 0.16 | 0.15 | 0.17 | 1.07 | 0.14* | 0.17 | 0.18 |
| $\sigma_{33}$ | 1 | 1.08 | 0.20 | 0.11 | 0.17 | 1.07 | 0.18 | 0.12 | 0.15 | 1.05 | 0.17 | 0.13 | 0.14 |
| $\sigma_{44}$ | 1 | 1.07 | 0.19 | 0.10 | 0.13 | 1.09 | 0.18* | 0.11 | 0.13 | 1.11 | 0.18* | 0.13 | 0.17 |
| $\sigma_{21}$ | 0.6 | 0.57 | 0.13 | 0.09 | 0.12 | 0.58 | 0.12 | 0.10 | 0.13 | 0.59 | 0.13 | 0.11 | 0.13 |
| $\sigma_{32}$ | 0.6 | 0.61 | 0.12 | 0.08 | 0.14 | 0.62 | 0.11 | 0.09 | 0.12 | 0.63 | 0.13 | 0.10 | 0.12 |
| $\sigma_{43}$ | 0 | 0.01 | 0.07 | 0.07 | 0.10 | 0.00 | 0.10 | 0.08 | 0.09 | -0.01 | 0.10 | 0.09 | 0.12 |
| $\sigma_{31}$ | 0 | -0.02 | 0.08 | 0.06 | 0.09 | -0.02 | 0.10 | 0.07 | 0.10 | 0.00 | 0.08 | 0.08 | 0.12 |
| $\sigma_{42}$ | 0 | -0.01 | 0.07 | 0.08 | 0.12 | 0.01 | 0.09 | 0.08 | 0.11 | 0.00 | 0.10 | 0.09 | 0.13 |
| $\sigma_{41}$ | 0.6 | 0.63 | 0.16 | 0.07 | 0.10 | 0.66 | 0.16 | 0.08 | 0.11 | 0.65 | 0.14 | 0.09 | 0.17 |

Note: $\sigma_{ii}$ denotes variance of $\beta_i$ and $\sigma_{ij}$ denotes covariance of $\beta_i$ and $\beta_j$.

**Table 4: Monte Carlo Results for G-MNL**

| | | Complete choiceset (60 choices) | | | | Random subset (20 choices) | | | | Random subset (10 choices) | | | |
| | | mean | empi. | mean ASE | | mean | empi. | mean ASE | | mean | empi. | mean ASE | |
| | TRUE | est | s.e. | BHHH | Robust | est | s.e. | BHHH | Robust | est | s.e. | BHHH | Robust |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta_1$ | 1 | **1.02** | 0.14 | 0.08 | 0.10 | **1.02** | 0.12 | 0.08 | 0.11 | **1.03** | 0.10 | 0.09 | 0.11 |
| $\beta_2$ | 1 | **1.01** | 0.10 | 0.07 | 0.10 | **1.02** | 0.09 | 0.08 | 0.12 | **1.03** | 0.09 | 0.09 | 0.11 |
| $\beta_3$ | 1 | **1.00** | 0.10 | 0.06 | 0.11 | **1.01** | 0.10 | 0.07 | 0.11 | **1.04** | 0.08* | 0.08 | 0.10 |
| $\beta_4$ | -1 | **-1.00** | 0.14 | 0.07 | 0.10 | **-1.01** | 0.11 | 0.07 | 0.10 | **-1.02** | 0.12 | 0.08 | 0.10 |
| | | | | | | | | | | | | | |
| $\sigma_{11}$ | 1 | **1.04** | 0.16 | 0.15 | 0.20 | **1.02** | 0.20 | 0.16 | 0.20 | **1.06** | 0.23 | 0.19 | 0.20 |
| $\sigma_{22}$ | 1 | **1.03** | 0.20 | 0.16 | 0.19 | **0.96** | 0.22 | 0.16 | 0.18 | **0.99** | 0.14 | 0.18 | 0.20 |
| $\sigma_{33}$ | 1 | **1.08** | 0.17* | 0.12 | 0.20 | **1.02** | 0.19 | 0.13 | 0.16 | **1.02** | 0.20 | 0.14 | 0.17 |
| $\sigma_{44}$ | 1 | **1.05** | 0.19 | 0.11 | 0.16 | **1.02** | 0.16 | 0.12 | 0.15 | **1.06** | 0.15 | 0.14 | 0.17 |
| $\sigma_{21}$ | 0.6 | **0.64** | 0.14 | 0.09 | 0.15 | **0.59** | 0.17 | 0.10 | 0.16 | **0.57** | 0.13 | 0.11 | 0.15 |
| $\sigma_{32}$ | 0.6 | **0.64** | 0.16 | 0.09 | 0.14 | **0.55** | 0.22 | 0.09 | 0.13 | **0.57** | 0.17 | 0.11 | 0.14 |
| $\sigma_{43}$ | 0 | **0.04** | 0.10 | 0.08 | 0.13 | **0.05** | 0.08* | 0.08 | 0.12 | **0.02** | 0.07 | 0.09 | 0.12 |
| $\sigma_{31}$ | 0 | **0.07** | 0.11* | 0.07 | 0.13 | **0.02** | 0.13 | 0.07 | 0.12 | **-0.03** | 0.13 | 0.08 | 0.12 |
| $\sigma_{42}$ | 0 | **0.02** | 0.16 | 0.08 | 0.12 | **0.03** | 0.11 | 0.08 | 0.12 | **0.04** | 0.10 | 0.10 | 0.11 |
| $\sigma_{41}$ | 0.6 | **0.61** | 0.18 | 0.07 | 0.13 | **0.64** | 0.16 | 0.09 | 0.13 | **0.66** | 0.12* | 0.10 | 0.13 |
| $\tau$ | 0.5 | **0.48** | 0.14 | 0.05 | 0.10 | **0.51** | 0.12 | 0.06 | 0.09 | **0.52** | 0.10 | 0.07 | 0.09 |
| $\gamma$ | 0.5 | **0.47** | 0.35 | 0.12 | 0.18 | **0.60** | 0.29 | 0.12 | 0.16 | **0.58** | 0.25 | 0.13 | 0.14 |

Note: $\sigma_{ii}$ denotes variance of $\beta_i$ and $\sigma_{ij}$ denotes covariance of $\beta_i$ and $\beta_j$.

**Table 5: Monte Carlo Results for MM-MNL**

| | TRUE | Complete choiceset (60 choices) | | mean ASE | | Random subset (20 choices) | | mean ASE | | Random subset (10 choices) | | mean ASE | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | mean est | empi. s.e. | BHHH | Robust | mean est | empi. s.e. | BHHH | Robust | mean est | empi. s.e. | BHHH | Robust |
| **class 1** | | | | | | | | | | | | | |
| $\beta_{11}$ | 1 | **0.98** | 0.13 | 0.10 | 0.13 | **0.98** | 0.14 | 0.11 | 0.14 | **0.98** | 0.16 | 0.12 | 0.15 |
| $\beta_{21}$ | 1 | **1.00** | 0.13 | 0.10 | 0.14 | **1.02** | 0.14 | 0.11 | 0.17 | **1.02** | 0.16 | 0.11 | 0.16 |
| $\beta_{31}$ | 1 | **1.01** | 0.15 | 0.07 | 0.12 | **1.02** | 0.14 | 0.08 | 0.11 | **1.04** | 0.15 | 0.09 | 0.12 |
| $\beta_{41}$ | -1 | **-0.97** | 0.14 | 0.07 | 0.13 | **-0.97** | 0.16 | 0.08 | 0.12 | **-0.97** | 0.15 | 0.09 | 0.13 |
| $\sqrt{\sigma_{11,1}}$ | 1 | **1.01** | 0.16 | 0.10 | 0.12 | **1.02** | 0.19 | 0.11 | 0.12 | **1.04** | 0.15 | 0.12 | 0.12 |
| $\sqrt{\sigma_{22,1}}$ | 1 | **0.94** | 0.15 | 0.10 | 0.12 | **0.94** | 0.13* | 0.11 | 0.16 | **0.96** | 0.16 | 0.11 | 0.15 |
| $\sqrt{\sigma_{33,1}}$ | 1 | **0.99** | 0.11 | 0.06 | 0.09 | **1.00** | 0.09 | 0.07 | 0.09 | **1.01** | 0.10 | 0.08 | 0.11 |
| $\sqrt{\sigma_{44,1}}$ | 1 | **0.99** | 0.10 | 0.06 | 0.10 | **0.98** | 0.11 | 0.06 | 0.10 | **1.03** | 0.09 | 0.07 | 0.10 |
| **class 2** | | | | | | | | | | | | | |
| $\beta_{12}$ | -1 | **-0.95** | 0.15 | 0.12 | 0.15 | **-0.92** | 0.19 | 0.13 | 0.17 | **-0.95** | 0.17 | 0.14 | 0.18 |
| $\beta_{22}$ | -1 | **-1.00** | 0.13 | 0.10 | 0.12 | **-1.02** | 0.15 | 0.11 | 0.13 | **-1.02** | 0.15 | 0.13 | 0.14 |
| $\beta_{32}$ | 1 | **0.98** | 0.13 | 0.11 | 0.14 | **0.99** | 0.12 | 0.12 | 0.14 | **0.99** | 0.16 | 0.13 | 0.15 |
| $\beta_{42}$ | -1 | **-1.01** | 0.18 | 0.11 | 0.13 | **-0.99** | 0.17 | 0.12 | 0.14 | **-1.02** | 0.19 | 0.13 | 0.15 |
| $\sqrt{\sigma_{11,2}}$ | 0.6 | **0.65** | 0.28 | 0.15 | 0.17 | **0.65** | 0.28 | 0.15 | 0.19 | **0.62** | 0.34 | 0.18 | 0.24 |
| $\sqrt{\sigma_{22,2}}$ | 0.6 | **0.49** | 0.26 | 0.13 | 0.13 | **0.47** | 0.29 | 0.14 | 0.14 | **0.51** | 0.17* | 0.18 | 0.19 |
| $\sqrt{\sigma_{33,2}}$ | 1 | **1.02** | 0.15 | 0.10 | 0.11 | **1.01** | 0.15 | 0.11 | 0.13 | **1.02** | 0.15 | 0.12 | 0.14 |
| $\sqrt{\sigma_{44,2}}$ | 1 | **1.04** | 0.15 | 0.08 | 0.11 | **1.05** | 0.13 | 0.09 | 0.12 | **1.03** | 0.16 | 0.11 | 0.12 |
| class prob. | **0.64** | **0.62** | 0.04 | 0.05 | 0.05 | **0.62** | 0.05 | 0.05 | 0.05 | **0.63** | 0.05 | 0.06 | 0.06 |

Note: $\sqrt{\sigma_{ii,s}}$ denotes standard deviation of $\beta_{is}$.

**Table 6:  The Bias in Information Criteria when Using a Subset of Alternatives**

| N-MIXL | Average LL | | LL$_{\text{N-MIXL}}$-LL$_{\text{MNL}}$ | Average BIC gain |
| --- | --- | --- | --- | --- |
| | MNL | N-MIXL | | |
| Choiceset size 60 | -13336 | -12953 | 383 | 741 |
| Choiceset size 20 | -9113 | -8778 | 335 | 646 |
| Choiceset size 10 | -6602 | -6323 | 279 | 533 |

| MM-MNL | Average LL | | LL$_{\text{MM-MNL}}$-LL$_{\text{MNL}}$ | Average BIC gain |
| --- | --- | --- | --- | --- |
| | MNL | MM-MNL | | |
| Choiceset size 60 | -14814 | -12741 | 2073 | 4037 |
| Choiceset size 20 | -10502 | -8633 | 1868 | 3629 |
| Choiceset size 10 | -7846 | -6242 | 1604 | 3100 |

Note: N-MIXL and MM-MNL use datasets from Table 2 and Table 5, respectively.

## Table 7: Estimates from Real Data (Based on a Random Subset of 20 Choices)

| | MNL | | S-MNL[a] | | N-MIXL[b] | | G-MNL[b] | | Latent class[c] | | | | | | MM-MNL[d] | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | | | class 1 | | class 2 | | class 3 | | class 1 | | class 2 | |
| | est | s.e. | est | s.e. | est | s.e. | est | s.e. | est. | s.e. | est. | s.e. | est. | s.e. | est. | s.e. | est. | s.e. |
| **Brand [omitted others]** | | | | | | | | | | | | | | | | | | |
| Tombstone | **0.83** | 0.05 | **0.65** | 0.24 | **0.48** | 0.12 | **0.57** | 0.12 | **2.19** | 0.39 | 0.23 | 0.53 | 0.30 | 0.29 | **1.44** | 0.18 | -0.35 | 0.50 |
| Roma | **0.24** | 0.05 | -0.26 | 0.16 | **-0.33** | 0.12 | -0.20 | 0.11 | -0.18 | 0.32 | 0.19 | 0.27 | **-1.48** | 0.47 | 0.10 | 0.19 | **-0.86** | 0.42 |
| Jacks | **0.63** | 0.05 | 0.23 | 0.25 | 0.22 | 0.12 | **0.27** | 0.11 | **2.53** | 0.25 | **0.83** | 0.34 | **-1.15** | 0.51 | **1.42** | 0.16 | **-2.07** | 0.46 |
| Red Baron | 0.11 | 0.06 | -0.08 | 0.20 | -0.13 | 0.11 | -0.16 | 0.11 | 0.32 | 0.36 | 0.58 | 0.31 | -0.04 | 0.24 | **0.45** | 0.15 | **-0.82** | 0.35 |
| Bernatello | **-1.03** | 0.06 | **-0.99** | 0.16 | **-1.01** | 0.09 | **-1.07** | 0.12 | -0.47 | 0.25 | -0.72 | 0.40 | -1.67 | 0.94 | **-0.44** | 0.10 | **-2.01** | 0.39 |
| Price | **-0.82** | 0.04 | **-1.02** | 0.09 | **-1.22** | 0.07 | **-1.23** | 0.07 | **-1.89** | 0.19 | **-1.66** | 0.23 | **-1.04** | 0.14 | **-1.62** | 0.11 | -0.26 | 0.14 |
| Promotion | **0.81** | 0.10 | **0.85** | 0.16 | **0.91** | 0.20 | **0.88** | 0.19 | **1.15** | 0.26 | **1.19** | 0.16 | 0.88 | 0.45 | **1.08** | 0.29 | 0.78 | 0.52 |
| **Toppings [omitted combo]** | | | | | | | | | | | | | | | | | | |
| Cheese only | **0.15** | 0.05 | -0.04 | 0.05 | -0.29 | 0.16 | **-0.41** | 0.17 | -0.09 | 0.48 | 0.47 | 0.46 | 0.15 | 0.64 | -0.24 | 0.22 | **-1.31** | 0.52 |
| Sausage/pepperoni | **1.02** | 0.04 | **0.89** | 0.08 | **0.99** | 0.09 | **1.00** | 0.07 | **1.13** | 0.19 | **0.82** | 0.40 | 0.30 | 0.47 | **0.71** | 0.13 | **1.37** | 0.30 |
| Meat/supreme | -0.06 | 0.05 | -0.07 | 0.05 | -0.12 | 0.09 | -0.12 | 0.09 | 0.12 | 0.24 | **-0.48** | 0.18 | 0.29 | 0.37 | -0.22 | 0.15 | -0.05 | 0.23 |
| Bacon/Burger | **-0.34** | 0.06 | **-0.46** | 0.07 | **-0.67** | 0.11 | **-0.75** | 0.15 | -0.21 | 0.23 | -0.96 | 0.52 | -0.35 | 0.70 | **-0.40** | 0.15 | -0.44 | 0.41 |
| Chicken/Mexican | **-0.51** | 0.08 | **-0.79** | 0.12 | **-0.78** | 0.11 | **-0.85** | 0.15 | **-0.91** | 0.31 | **-1.55** | 0.62 | 0.01 | 0.64 | **-0.74** | 0.18 | -0.26 | 0.30 |
| Vegetarian | **-0.99** | 0.15 | **-1.71** | 0.22 | **-1.40** | 0.23 | **-2.07** | 0.35 | **-2.65** | 1.08 | **-2.66** | 0.96 | 0.13 | 0.87 | **-2.55** | 0.44 | -0.43 | 0.63 |
| **Crust [omitted regular, others]** | | | | | | | | | | | | | | | | | | |
| Rising | **-0.84** | 0.04 | **-1.04** | 0.09 | **-1.20** | 0.12 | **-1.13** | 0.11 | **-1.50** | 0.37 | **-0.90** | 0.32 | -0.29 | 0.36 | **-1.88** | 0.18 | **-0.81** | 0.25 |
| Thin/crispy | 0.03 | 0.03 | 0.01 | 0.03 | -0.02 | 0.06 | -0.03 | 0.05 | **-0.41** | 0.14 | 0.09 | 0.30 | **-0.57** | 0.24 | 0.06 | 0.09 | -0.22 | 0.17 |
| Microwavable | 0.03 | 0.06 | **0.14** | 0.05 | -0.06 | 0.08 | -0.06 | 0.10 | 0.26 | 0.61 | **0.84** | 0.32 | -1.06 | 1.07 | 0.29 | 0.19 | -0.73 | 0.39 |
| $\tau$ | | | **0.92** | 0.07 | | | **0.37** | 0.04 | | | | | | | | | | |
| $\gamma$ | | | | | | | **0.80** | 0.10 | | | | | | | | | | |
| class prob. | | | | | | | | | **0.30** | 0.06 | **0.25** | 0.08 | **0.19** | 0.06 | **0.73** | 0.08 | **0.27** | 0.08 |
| | 16 | | 27 | | 48 | | 50 | | 84 | | | | | | 66 | | | |
| LL | -10814 | | -9285 | | -8538 | | -8524 | | -9417 | | | | | | **-8343** | | | |
| BIC | 21761 | | 18795 | | 17476 | | 17464 | | 19533 | | | | | | **17234** | | | |

Note: [a] estimates from S-MNL with random correlated (one-factor) intercepts; [b] estimates from correlated coefficients (imposing 1-factor structure on the covariance matrix); [c] estimates from LC with 5 classes; [d] estimates from MM-MNL with 2 proportional covariance matrices. Bold estimates are statistically significant at 5%. S-MNL, N-MIXL, G-MNL and MM-MNL are estimated by simulated maximum likelihood with 500 draws. The standard errors are calculated using 5000 draws.

**Table 8: Estimates from Real Data (Based on a Random Subset of 30 Choices)**

| | MNL | | S-MNL[a] | | N-MIXL[b] | | G-MNL[b] | | Latent class[c] | | | | | | MM-MNL[d] | | | |
| | | | | | | | | | class 1 | | class 2 | | class 3 | | class 1 | | class 2 | |
| | est | s.e. | est | s.e. | est | s.e. | est | s.e. | est. | s.e. | est. | s.e. | est. | s.e. | est. | s.e. | est. | s.e. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Brand [omitted others] | | | | | | | | | | | | | | | | | | |
| Tombstone | **0.86** | 0.05 | **0.71** | 0.21 | **0.40** | 0.11 | **0.48** | 0.11 | 0.87 | 1.32 | **2.64** | 0.47 | 0.08 | 0.36 | **1.63** | 0.16 | -0.34 | 0.31 |
| Roma | **0.23** | 0.05 | -0.21 | 0.17 | **-0.30** | 0.10 | **-0.29** | 0.10 | -0.25 | 0.28 | -0.61 | 0.64 | **1.51** | 0.34 | -0.05 | 0.19 | **-0.74** | 0.21 |
| Jacks | **0.64** | 0.05 | 0.26 | 0.22 | **0.20** | 0.10 | 0.06 | 0.12 | 0.43 | 0.40 | **2.67** | 0.87 | -0.03 | 0.44 | **1.47** | 0.14 | **-2.12** | 0.37 |
| Red Baron | **0.12** | 0.06 | -0.07 | 0.21 | **-0.25** | 0.11 | **-0.34** | 0.12 | 0.48 | 0.48 | 0.61 | 0.72 | 0.20 | 0.33 | **0.47** | 0.14 | **-0.83** | 0.30 |
| Bernatello | **-1.03** | 0.06 | **-0.95** | 0.15 | **-0.99** | 0.09 | **-1.07** | 0.10 | **-0.92** | 0.29 | -0.51 | 0.51 | **-0.98** | 0.41 | **-0.42** | 0.11 | **-1.81** | 0.34 |
| Price | **-0.83** | 0.04 | **-1.01** | 0.07 | **-1.18** | 0.06 | **-1.16** | 0.06 | **-1.62** | 0.44 | **-1.85** | 0.30 | **-0.80** | 0.18 | **-1.66** | 0.11 | **-0.33** | 0.11 |
| Promotion | **0.83** | 0.10 | **0.84** | 0.15 | **0.88** | 0.18 | **0.79** | 0.18 | **1.16** | 0.24 | **0.98** | 0.34 | **0.94** | 0.18 | **1.05** | 0.27 | 0.63 | 0.50 |
| Toppings [omitted combo] | | | | | | | | | | | | | | | | | | |
| Cheese only | **0.16** | 0.05 | -0.03 | 0.05 | **-0.32** | 0.15 | **-0.31** | 0.15 | -0.54 | 0.41 | -0.03 | 0.34 | 0.28 | 0.42 | -0.32 | 0.19 | **-1.09** | 0.35 |
| Sausage/pepperoni | **1.02** | 0.04 | **0.88** | 0.07 | **0.99** | 0.07 | **1.03** | 0.07 | **0.58** | 0.17 | **1.21** | 0.31 | **1.24** | 0.25 | **0.73** | 0.11 | **1.45** | 0.26 |
| Meat/supreme | -0.06 | 0.05 | -0.07 | 0.05 | -0.002 | 0.08 | 0.02 | 0.08 | 0.07 | 0.78 | 0.10 | 0.91 | -0.09 | 0.21 | -0.20 | 0.13 | -0.13 | 0.22 |
| Bacon/burger | **-0.36** | 0.06 | **-0.47** | 0.06 | **-0.60** | 0.10 | **-0.72** | 0.14 | -0.31 | 0.23 | -0.46 | 0.30 | -0.13 | 0.31 | **-0.40** | 0.16 | -0.49 | 0.31 |
| Chicken/Mexican | **-0.52** | 0.08 | **-0.79** | 0.10 | **-0.61** | 0.11 | **-0.80** | 0.17 | -0.29 | 0.25 | **-1.19** | 0.46 | -0.01 | 0.91 | **-0.67** | 0.17 | -0.20 | 0.33 |
| Vegetarian | **-1.04** | 0.15 | **-1.70** | 0.18 | **-1.56** | 0.26 | **-2.13** | 0.35 | -1.43 | 2.09 | -2.93 | 1.67 | -0.66 | 0.80 | **-2.73** | 0.42 | -0.34 | 0.42 |
| Crust [omitted regular, others] | | | | | | | | | | | | | | | | | | |
| Rising | **-0.85** | 0.04 | **-1.03** | 0.07 | **-1.13** | 0.12 | **-1.03** | 0.10 | **-1.10** | 0.37 | **-1.62** | 0.33 | **-1.17** | 0.34 | **-1.83** | 0.14 | **-0.66** | 0.21 |
| Thin/crispy | 0.03 | 0.03 | 0.02 | 0.02 | -0.03 | 0.06 | -0.06 | 0.05 | -0.14 | 0.63 | -0.67 | 0.28 | **0.98** | 0.24 | 0.02 | 0.09 | -0.12 | 0.15 |
| Microwavable | 0.05 | 0.06 | **0.17** | 0.05 | -0.08 | 0.09 | -0.07 | 0.10 | 0.12 | 0.61 | 0.52 | 0.82 | 0.02 | 0.32 | **0.46** | 0.18 | **-0.70** | 0.30 |
| $\tau$ | | | **0.88** | 0.06 | | | **0.43** | 0.04 | | | | | | | | | | |
| $\gamma$ | | | | | | | **0.59** | 0.07 | | | | | | | | | | |
| class prob. | | | | | | | | | **0.33** | 0.15 | 0.26 | 0.17 | **0.20** | 0.05 | **0.72** | 0.08 | **0.28** | 0.08 |
| | 16 | | 27 | | 48 | | 50 | | 84 | | | | | | 66 | | | |
| LL | -12441 | | -10820 | | -9969 | | -9944 | | -10933 | | | | | | **-9751** | | | |
| BIC | 25015 | | 21865 | | 20338 | | 20304 | | 22565 | | | | | | **20052** | | | |

Note: [a] estimates from S-MNL with random correlated (one-factor) intercepts; [b] estimates from correlated coefficients (imposing 1-factor structure on the covariance matrix); [c] estimates from LC with 5 classes; [d] estimates from MM-MNL with 2 proportional covariance matrices. Bold estimates are statistically significant at 5%. S-MNL, N-MIXL, G-MNL and MM-MNL are estimated by simulated maximum likelihood with 500 draws. The standard errors are calculated using 5000 draws.

## Table 9: Estimates from Real Data (Based on a Random Subset of 40 Choices)

| | MNL | | S-MNL[a] | | N-MIXL[b] | | G-MNL[b] | | Latent class[c] | | | | | | MM-MNL[d] | | | |
| | | | | | | | | | class 1 | | class 2 | | class 3 | | class 1 | | class 2 | |
| | est | s.e. | est | s.e. | est | s.e. | est | s.e. | est. | s.e. | est. | s.e. | est. | s.e. | est. | s.e. | est. | s.e. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Brand [omitted others]** | | | | | | | | | | | | | | | | | | |
| Tombstone | **0.87** | 0.05 | **0.68** | 0.21 | **0.52** | 0.10 | **0.48** | 0.11 | **2.04** | 0.31 | **0.64** | 0.31 | 1.12 | 0.59 | **1.71** | 0.16 | -0.45 | 0.23 |
| Roma | **0.24** | 0.05 | -0.25 | 0.17 | **-0.26** | 0.11 | **-0.34** | 0.10 | **-0.69** | 0.30 | **0.91** | 0.36 | -0.89 | 1.20 | -0.30 | 0.17 | **-0.58** | 0.25 |
| Jacks | **0.66** | 0.05 | 0.25 | 0.21 | **0.29** | 0.10 | 0.01 | 0.12 | **2.30** | 0.32 | 0.17 | 0.36 | 0.24 | 0.47 | **1.52** | 0.18 | **-1.98** | 0.30 |
| Red Baron | **0.13** | 0.06 | -0.12 | 0.20 | -0.17 | 0.12 | -0.17 | 0.10 | -0.30 | 0.28 | 0.37 | 0.33 | **1.09** | 0.28 | **0.55** | 0.17 | -0.43 | 0.23 |
| Bernatello | **-1.02** | 0.06 | **-0.98** | 0.15 | **-0.98** | 0.08 | **-1.08** | 0.10 | **-0.68** | 0.30 | **-1.17** | 0.27 | **-1.40** | 0.50 | **-0.57** | 0.10 | **-1.35** | 0.13 |
| Price | **-0.84** | 0.04 | **-0.97** | 0.06 | **-1.21** | 0.06 | **-1.20** | 0.06 | **-1.85** | 0.14 | **-1.65** | 0.24 | **-1.07** | 0.15 | **-1.75** | 0.07 | **-0.48** | 0.09 |
| Promotion | **0.81** | 0.10 | **0.77** | 0.13 | **0.86** | 0.17 | **0.76** | 0.15 | **1.11** | 0.22 | **1.18** | 0.21 | -0.04 | 0.79 | **0.92** | 0.23 | 0.65 | 0.41 |
| **Toppings [omitted combo, others]** | | | | | | | | | | | | | | | | | | |
| Cheese only | **0.16** | 0.05 | -0.01 | 0.05 | -0.28 | 0.13 | **-0.42** | 0.14 | -0.38 | 0.29 | **-1.34** | 0.39 | 0.52 | 0.28 | -0.24 | 0.17 | **-1.01** | 0.30 |
| Sausage/pepperoni | **1.02** | 0.04 | **0.85** | 0.07 | **1.02** | 0.07 | **1.04** | 0.06 | **1.17** | 0.18 | **1.14** | 0.19 | 0.42 | 0.27 | **0.61** | 0.10 | **1.53** | 0.21 |
| Meat/supreme | -0.06 | 0.05 | -0.08 | 0.04 | -0.04 | 0.08 | -0.05 | 0.07 | 0.27 | 0.21 | -0.05 | 0.16 | -0.43 | 0.37 | **-0.38** | 0.12 | -0.26 | 0.26 |
| Bacon/burger | **-0.36** | 0.06 | **-0.44** | 0.06 | **-0.64** | 0.10 | **-0.83** | 0.12 | -0.21 | 0.24 | -0.31 | 0.25 | -0.89 | 0.50 | **-0.62** | 0.15 | -0.22 | 0.27 |
| Chicken/Mexican | **-0.53** | 0.08 | **-0.76** | 0.09 | **-0.71** | 0.10 | **-0.89** | 0.14 | **-1.02** | 0.23 | 0.08 | 0.53 | -0.55 | 0.54 | **-1.19** | 0.22 | 0.18 | 0.22 |
| Vegetarian | **-1.07** | 0.15 | **-1.66** | 0.17 | **-1.75** | 0.23 | **-2.19** | 0.34 | **-3.08** | 0.92 | **-1.74** | 0.63 | -0.42 | 0.76 | **-3.76** | 0.45 | -0.18 | 0.37 |
| **Crust [omitted regular, others]** | | | | | | | | | | | | | | | | | | |
| Rising | **-0.85** | 0.04 | **-1.00** | 0.06 | **-1.16** | 0.08 | **-1.11** | 0.10 | **-1.45** | 0.26 | **-1.08** | 0.23 | **-1.10** | 0.23 | **-1.62** | 0.14 | **-1.04** | 0.19 |
| Thin/crispy | 0.03 | 0.03 | 0.02 | 0.02 | -0.05 | 0.07 | -0.01 | 0.05 | **-0.62** | 0.12 | **0.74** | 0.18 | -0.46 | 0.79 | **-0.32** | 0.11 | 0.24 | 0.17 |
| Microwavable | 0.07 | 0.06 | **0.16** | 0.04 | 0.02 | 0.08 | 0.01 | 0.10 | 0.64 | 0.37 | **0.90** | 0.26 | -1.04 | 1.51 | **0.67** | 0.16 | **-0.89** | 0.23 |
| $\tau$ | | | 0.86 | 0.05 | | | **0.40** | 0.04 | | | | | | | | | | |
| $\gamma$ | | | | | | | **0.79** | 0.07 | | | | | | | | | | |
| class prob. | | | | | | | | | **0.28** | 0.04 | **0.24** | 0.05 | 0.23 | 0.15 | **0.65** | 0.08 | **0.35** | 0.08 |
| | 16 | | 27 | | 48 | | 50 | | 84 | | | | | | 66 | | | |
| LL | -13602 | | -11930 | | -11048 | | -11017 | | -12116 | | | | | | **-10802** | | | |
| BIC | 27337 | | 24085 | | 22497 | | 22450 | | 24931 | | | | | | **22153** | | | |

Note: [a] estimates from S-MNL with random correlated (one-factor) intercepts; [b] estimates from correlated coefficients (imposing 1-factor structure on the covariance matrix); [c] estimates from LC with 5 classes; [d] estimates from MM-MNL with 2 proportional covariance matrices. Bold estimates are statistically significant at 5%. S-MNL, N-MIXL, G-MNL and MM-MNL are estimated by simulated maximum likelihood with 500 draws. The standard errors are calculated using 5000 draws.

# Table 10: MM-MNL Model Fit to Market Shares by Brand/Topping

**Observed Market Shares**

|  | cheese | sausage/ pepperoni | meat/ supreme | bacon/ burger | chicken/ mexican | Veg | Other | Total |
|---|---|---|---|---|---|---|---|---|
| Tombstone | 1.71 | 11.03 | 4.21 | 2.48 | 1.76 | 0.33 | 4.79 | **26** |
| Roma | 3.76 | 10.97 | 1.95 | 0.69 | ----- | ----- | 2.84 | **20** |
| Jacks | 2.26 | 10.18 | 2.59 | 3.43 | 0.70 | ----- | 1.31 | **20** |
| Red Baron | 2.71 | 4.54 | 2.34 | 0.47 | 0.45 | ----- | 0.09 | **11** |
| Bernatello | 1.14 | 2.85 | 0.31 | 0.58 | 0.03 | ----- | 0.86 | **6** |
| Others | 2.45 | 9.07 | 2.46 | 0.02 | 0.22 | 0.91 | 1.50 | **17** |
| **Total** | **14** | **49** | **14** | **8** | **3** | **1** | **11** | **100** |

**Predicted Market Shares by MM-MNL Model**

|  | cheese | sausage/ pepperoni | meat/ supreme | bacon/ burger | chicken/ mexican | Veg | Other | Total |
|---|---|---|---|---|---|---|---|---|
| Tombstone | 2.56 | 11.55 | 4.02 | 3.03 | 0.89 | 0.66 | 3.49 | **26** |
| Roma | 3.86 | 11.48 | 1.49 | 0.48 | ----- | ----- | 2.72 | **20** |
| Jacks | 2.17 | 10.08 | 2.93 | 2.87 | 0.96 | ----- | 1.45 | **20** |
| Red Baron | 2.34 | 4.42 | 2.71 | 0.64 | 0.54 | ----- | 0.11 | **11** |
| Bernatello | 0.73 | 2.81 | 0.42 | 0.72 | 0.11 | ----- | 1.07 | **6** |
| Others | 2.59 | 7.96 | 2.25 | 0.04 | 0.73 | 0.50 | 2.61 | **17** |
| **Total** | **14** | **48** | **14** | **8** | **3** | **1** | **11** | **100** |

Note: There are seven brand/topping combinations that do not exist in the data, leaving 37 options.

# Table 11: Brand Level Price Elasticities Across Models and Choice Set Sizes

|  | Random Choice Set Size for Estimation | MNL | S-MNL | LC | MXL | G-MNL | MM-MNL |
|---|---|---|---|---|---|---|---|
| Tombstone | 20 | -1.66 | -1.49 | -1.75 | -1.46 | -1.40 | -1.69 |
|  | 30 | -1.69 | -1.48 | -2.05 | -1.49 | -1.49 | -1.72 |
|  | 40 | -1.71 | -1.50 | -2.09 | -1.52 | -1.42 | -1.66 |
| Jacks | 20 | -1.64 | -1.52 | -2.44 | -1.81 | -1.63 | -2.06 |
|  | 30 | -1.67 | -1.55 | -2.45 | -1.82 | -1.69 | -2.05 |
|  | 40 | -1.69 | -1.55 | -2.37 | -1.84 | -1.66 | -2.14 |

# Table 12: Decomposition of Variety Level Price Elasticities

### A. Tombstone sausage/pepperoni (Share = 11%)

|  | Change in Market Share (percentage points) | Distribution of Switchers | Decomposition of elasticity |
|---|---|---|---|
| Tombstone Sausage/Pepperoni | -1.81 |  | -1.68 |
| Tombstone Other Varieties | +0.48 | 27% | 0.45 |
| Other Brands Sausage/Pepperoni | +0.75 | 41% | 0.69 |
| Other Brands Other Varieties | +0.58 | 32% | 0.54 |

### B. Jacks meat/supreme (Share = 2.9%)

|  | Change in Market Share (percentage points) | Distribution of Switchers | Decomposition of elasticity |
|---|---|---|---|
| Jacks Meat/Supreme | -0.97 |  | -3.34 |
| Jacks Other Varieties | +0.41 | 43% | 1.43 |
| Other Brands Meat/Supreme | +0.13 | 14% | 0.46 |
| Other Brands Other Varieties | +0.42 | 43% | 1.45 |