The Effects of Parameterization on Heterogeneous Choice Models

Garrett Sonnier, Andrew Ainslie, and Thomas Otter[1]

DRAFT: Please do not circulate or cite without permission from the authors.

[1] Sonnier (garrett.sonnier@anderson.ucla.edu) is a Ph.D. student and Ainslie (andrew.ainslie@anderson.ucla.edu) an Assistant Professor at UCLA, 110 Westwood Plaza, Los Angeles, CA 90095, 310-2068585, fax 310-2067422. Otter (otter_2@cob.osu.edu) is an Assistant Professor at The Ohio State University, 2100 Neil Avenue, Columbus, OH 43210, 614-2476433. The authors would like to thank Jordan Louviere, Kenneth Train, and Greg Allenby for helpful insights. Seminar participants at The Ohio State University also provided useful comments.

The Effects of Parameterization on Heterogeneous Choice Models

ABSTRACT:
It is well-known that the choice probabilities derived from random utility models suffer from scale invariance. The usual approach to this identification problem is to place some type of restriction on the parameter space. Generally, the specific restriction employed has been regarded as an arbitrary choice. In models with homogeneous coefficients or fixed effect approaches to heterogeneity, the model fits are unaffected and the parameters estimated under one type of restriction can easily be recovered by transforming those from another restriction. This typically leads researchers to choose the parameterization that most easily facilitates estimation, even though this specification leads to parameters with ambiguous interpretability. However, we demonstrate that the choice of parameterization in a Bayesian setting that incorporates prior distributions of heterogeneity is not arbitrary. We show that the model fits are different and the duality of parameter estimates under alternative restrictions is not generally maintained. We illustrate these issues with two specifications of choice models prevalently employed in the literature.

1. Introduction

Modeling consumer choice has long been a central focus of the marketing literature. The majority of empirical choice models in the literature are derived from the theory of random utility. It is well-known that the choice probabilities derived from this approach suffer from scale invariance. The usual approach to this identification problem is to place some type of restriction on the parameter space. With few exceptions, researchers have not paid much attention to the choice of restriction. It is regarded as an arbitrary choice; model fits are unaffected and parameters estimated under one type of restriction can easily be recovered by transforming those from another restriction. This typically leads researchers to choose the parameterization that most easily facilitates estimation, even if this specification results in parameters with ambiguous interpretability. To the extent that a different set of parameters are of interest, they are easily recovered through the appropriate transformation. This approach to parameterization is only justified, though, if the choice is truly arbitrary.

The conclusion that, empirically, model parameterization is an arbitrary choice is based on classical inference and models with no heterogeneity or discrete representations of heterogeneity. In a random coefficients setting, with distributions specified on heterogeneous model coefficients, the choice of parameterization is not at all arbitrary. Model fits can be affected and the duality of parameter estimates under different model specifications is not generally maintained. Thus, choice of parameterization in this setting can be influential. We illustrate the issues involved by examining two specifications of choice models prevalently employed in the literature. The first specifies parameters on product attributes and the logarithm of price, while the second specifies parameters on product attributes and price. In the former, the economic model is derived from maximizing utility per dollar (Arora et al. 1998). In the latter, the economic model is derived from either consumer surplus maximization or utility maximization (Hanemann and Kanninen 1998; Jedidi et al. 2003).

The main contribution of this paper is to demonstrate that alternative restrictions placed on choice models for scale identification do not always lead to the same estimates of structural parameters and fit to empirical data. Even when the underlying economic problems of the alternative specifications are dual, choice of parameterization can matter. Substantively, this is not merely a pedantic point regarding specification of choice models. The aforementioned conventional parameterizations of heterogeneous choice models are actually re-parameterizations of models grounded in micro-economic theory. In conjunction with standard priors, these re-parameterized models can lead to estimates of the underlying structural parameters that are severely skewed and prone to outliers. There is also no straightforward way to assess the effects of covariates on the structural parameters. Rather, the conventional parameterization tends to establish regressions on parameters with ambiguous economic meaning. Specifying the model in the space that directly identifies the structural parameters easily facilitates prior distributions that restrict outlier estimates and leads to more credible estimates. Investigating the impact of demographic covariates is readily handled in a system that properly accounts for uncertainty and is less susceptible to outliers.

The remainder of this paper is organized as follows. Section 2 briefly discusses some of the literature relevant to this topic. Section 3 discusses economic and statistical model specification. Section 4 presents applications of the concepts to two data sets:

(1) a choice-based conjoint data set on cameras and (2) a choice-based conjoint data set on mid-size sedans. Section 5 further investigates the properties of alternative model parameterizations with synthetic data. We summarize and conclude in Section 6.

2. Review of the Literature

Modeling heterogeneity has occupied a large space in the marketing literature on consumer choice. Many of the choice models that account for heterogeneity in response parameters attempt to relate variation in parameter estimates to household or individual characteristics (e.g. demographics, category purchase intensity). Rossi et al. (1996) and Ainslie and Rossi (1998) relate individual level brand intercepts and marketing mix parameter estimates to demographic covariates in a Hierarchical Bayes probit model of brand choice. Both of these hierarchical models regress the estimated coefficients of the latent utility regression onto demographic and other covariates. As noted by these authors, the economic meaning of these coefficients is somewhat ambiguous. The coefficients do not measure marginal utility, rather, marginal utility relative to the variance of the normally distributed error term (Swait and Louviere 1993).

Some authors have investigated alternative specifications of heterogeneous choice models that directly identify structural parameters of interest. Arora et al. (1998) specify and estimate a Bayesian model of primary and secondary demand. The consumer's discrete economic problem is specified as maximizing utility per dollar. Their model separately identifies the parameters that govern the marginal utility of changes in product attributes and μ, the scale parameter of the extreme value error term. Estimates of μ are crucial to linking the discrete and continuous parts of their model. The authors note, though, that the parameter $-\frac{1}{\mu}$ (specified on the log of price) is often interpreted as a reduced-form price coefficient after re-parameterization, such that $\alpha = -\frac{1}{\mu}$. Jedidi et al. (2003) specify a multivariate probit model of choice using consumer surplus maximization, which directly yields heterogeneous WTP estimates for products and product attributes. In their model, the coefficients on the price terms are set to the negative inverse of the standard deviations of the multivariate normal error distribution. Since the coefficient is on price rather than log of price, the non-price coefficients measure the WTP for changes in non-price attributes. It is straightforward to show that for models based on linear utility maximization, the WTP for changes in non-price attributes is the ratio of the non-price coefficient to the price coefficient (Train 2003).

In developing their model, Arora et al. (1998) first derive the economic and statistical specifications of their model without heterogeneity. Their discussion of parameterization uses the homogeneous model. In specifying their model, Jedidi et al. (2003) appeal to the duality of the economic problems of utility maximization and consumer surplus maximization (Hanemann and Kanninen 1998; Jedidi and Zhang 2002). Empirically, Cameron and James (1987) show that surplus maximization and utility maximization lead to the same WTP estimates and model fits with MLE estimation. The issue of empirical duality, though, has not been addressed in the context of random coefficient choice models. In this setting, the researcher must specify a prior distribution for the heterogeneous model parameters. If subsequently, the parameters are to be transformed to yield estimates of structural parameters of interest, standard priors can result in estimates with undesirable properties. Further, it is extremely difficult, if not impossible, to choose priors that will yield the same estimate of the structural parameter across different specifications. This issue is exacerbated by the fact that we typically have a relatively small number of observations per household with which to work.

3. Economic and Statistical Model Specification

Consider the consumer's discrete choice problem in Arora et al. (1998). From the set of M alternatives, the consumer chooses that good j which maximizes utility per dollar

(1) $$\frac{V_j}{p_j} \geq \frac{V_m}{p_m} \quad \forall m = 1, ..., M$$

where V_j is the utility of the j^{th} alternative and p_j is the price. Taking logs and expressing log marginal utility as a linear function of non-price attributes with an additive extreme value error specification yields a multinomial logit (MNL) choice probability for good j of

(2) $$\Pr[I = j] = \left[\frac{\exp\left[\dfrac{x_j^{'}\beta - \ln p_j}{\mu}\right]}{\sum_{m=1}^{M} \exp\left[\dfrac{x_m^{'}\beta - \ln p_m}{\mu}\right]} \right].$$

The structural parameters of interest are β, the parameters that govern the marginal utility of a change in x, and μ, the scale parameter of the extreme value error distribution. Equation (2) is more commonly re-parameterized such that $\varphi = \dfrac{\beta}{\mu}$ and $\gamma = -\dfrac{1}{\mu}$, which leads to

(3) $$\Pr[I=j]=\left[\frac{\exp\left[x_j'\varphi+\gamma\ln p_j\right]}{\sum_{m=1}^{M}\exp\left[x_m'\varphi+\gamma\ln p_m\right]}\right].$$

This linear re-parameterization facilitates estimation and, under certain circumstances, is costless. By the invariance property of the MLE estimator, maximizing the objective function by choosing $\left[\beta \quad \mu\right]'$ or $\left[\varphi \quad \gamma\right]'$ will yield the same fit to the observed data, the same estimates of the structural parameters (via the appropriate transformations), and hence the same the same estimate of the elasticities of the choice probabilities (Zehna 1966). In this sense, the choice of parameterization may be regarded as arbitrary.

An alternative view of the consumer's discrete choice problem is outlined in Jedidi et al. (2003). Here, the consumer is viewed as choosing the alternative that maximizes her surplus. Define the consumer's WTP for an alternative as a function of the alternative's non-price attributes and an additive extreme value error term. Consumer surplus is then defined as the difference between the consumer's WTP, or reservation price, and the market price. The consumer chooses the alternative j that yields the maximum surplus

(4) $$C_j - p_j \geq C_m - p_m \quad \forall m = 1,\ldots,M$$

where C_j is the WTP for alternative j and p_j is the price. The MNL choice probabilities take the form

(5) $$\Pr[I=j]=\left[\frac{\exp\left[\frac{x_j'\beta-p_j}{\mu}\right]}{\sum_{m=1}^{M}\exp\left[\frac{x_m'\beta-p_m}{\mu}\right]}\right].$$

Here, the structural parameters of interest are β, the willingness-to-pay for a change in the level of non-price attributes, and μ, the scale parameter. As before, the choice probabilities more commonly employed re-parameterize (5) such that $\varphi = \frac{\beta}{\mu}$ and $\gamma = -\frac{1}{\mu}$, which yields

$$(6) \qquad \Pr[I = j] = \left[\frac{\exp\left[x_j^{'}\phi + \gamma p_j \right]}{\sum_{m=1}^{M} \exp\left[x_m^{'}\phi + \gamma p_m \right]} \right].$$ [2]

Again, model fits from MLE estimation are unaffected, and estimates of WTP are available via $\beta = \frac{\varphi}{\gamma}$.

The re-parameterized model represented in (6) can be derived from the economic problem of maximizing a normalized linear indirect utility function, $v_j = x_j^{'}\varphi + \gamma \left(y - p_j \right) + \varepsilon_j$, where ε_j is distributed normalized extreme value. Starting from utility maximization, it is straightforward to compute what the consumer would be willing to pay for a change in x_{jk}, the k^{th} attribute of alternative j. It is the price change that would leave the individual indifferent between the alternative with the new level and the alternative with the original level. Assuming x_k is continuous, we take the total derivative of utility with respect to attribute k and price and set this to zero. Thus, we have $\partial v_j = \varphi_k \cdot \partial x_{jk} - \gamma \cdot \partial p_j = 0$ and the change in price that keeps utility constant given a

[2] For ease of exposition, we will refer to models parameterized as (2) and (5) as "non-linear" and the re-parameterized versions of these models as "linear".

change in attribute k is $\frac{\varphi_k}{\gamma}$ (Train 2003). Analogous arguments can be made for the case of categorical x's. Note that this quantity is directly available via β_k in equation (5).

For the models presented in (2) and (5), the choice of parameterization may be regarded as arbitrary. The fit of the model to the data is not affected and we can easily move back and forth between the alternative parameterizations with the appropriate identities. Tables 1a and 1b illustrates this point using our choice based conjoint data for a model without heterogeneity.[3]

---Insert Tables 1a and 1b Here---

Consider now a Bayesian approach to estimation of a heterogeneous choice model. We typically specify models in a hierarchical system where consumers are assumed to be heterogeneous in the parameters. There is a hierarchical prior for the consumer specific $\left[\beta_i' \quad \mu_i\right]'$ or $\left[\varphi_i' \quad \gamma_i\right]'$ where i indexes consumers. It is precisely the hierarchical prior that makes choice of parameterization influential in that different parameterizations in combination with standard hierarchical priors can lead to distinguishable models in terms of fit to empirical data and posterior inference. The shrinkage structure is placed on different parameters across the specifications, typically without regard to the mapping that links the structural parameters across specifications.

With the linear re-parameterizations, the researcher takes ratios of model coefficients to recover the structural parameters of interest. A standard hierarchical random coefficients model will introduce a parametric prior distribution for the non-price

[3] See Section 4 for a more detailed discussion of the data.

and price coefficients, $[\varphi_i \quad \gamma_i]' \sim G(\bullet)$. If $G(\bullet)$, which is typically the multivariate normal distribution, has positive density for combinations of $(\varphi_i \neq 0) \wedge (|\gamma_i| < \varepsilon)$ for small ε approaching zero, this prior readily allows for unbounded estimates of the structural parameters. To the extent that the posterior distribution $\pi(\varphi_i, \gamma_i \mid data)$ has positive density for such combinations, the posterior estimates of structural parameters will be unbounded. In addition to concerns about face validity, these outlier estimates can adversely impact the effort to assess the impact of covariates such as demographics on variation in structural parameters. Furthermore, if the prior for γ_i admits non-negative values, as does the MVN prior, the structural parameters can be undefined. This problem can be handled by choosing a prior that admits only negative values for the linear model price coefficient. However, it is generally not straightforward to incorporate knowledge about the structural parameters into $G(\bullet)$.

If the model is specified by (2) or (5), standard parametric prior distributions like the normal and log-normal for β and μ, respectively, can be used. While this parameterization more readily allows us to implement sensible prior structures on structural parameters, the parameterizations are not equivalent and the choice between them no longer arbitrary. We note that it is possible, in theory, to search for priors on $[\varphi_i' \quad \gamma_i]'$ that correspond to a normal prior for β_i and a log-normal prior for μ_i. This is not the goal of this paper, though. Rather, our goal is to ascertain what the linear re-parameterization in concert with standard priors implies for model fits and estimates of the underlying structural parameters. In the next section, we provide some empirical evidence on the effects of parameterization.

4. Two Choice-Based-Conjoint Applications

Choice-based conjoint (CBC) analysis is a key tool for marketing managers seeking to assess the market for new products. With the diffusion of HB estimation techniques that account for respondent heterogeneity, CBC has grown in popularity among the practitioner community. CBC readily lends itself to product volume estimates and estimates of respondent valuations of the product attributes. Using CBC data sets provided to us by firms in the camera and automotive categories, we will demonstrate the effects of parameterization on heterogeneous choice models. Table 2 presents the attributes and levels involved in the design of each study.

---Insert Table 2 Here---

Study 1: Cameras

The first data set is CBC data on cameras. The study was conducted by the Eastman Kodak Company to assess the market for a new camera format, the Advanced Photo System (APS). A detailed description of the data is given by Gilbride and Allenby (2003). A total of 302 respondents participated in the study. Each respondent completed 14 choice tasks, with each task consisting of three 35mm cameras, three APS cameras, and a no-buy option. Some attributes were available only on the APS camera, and price was nested within camera type. Across all the tasks, the average price for the 35MM cameras was between $80 and $170, while the price for the APS cameras was between $224 and $418. Given the design, we might expect price changes at lower prices to have more impact than at higher prices. Indeed, preliminary analyses revealed that models using log price fit the data much better than models using price. Thus, the CBC camera data is used to estimate the model based on maximization of utility per dollar. Log marginal utility is expressed as a linear function of the K non-price attributes. For identification, the lowest level of each attribute (with the exception of the body type) is dropped. This is the same coding scheme employed by Gilbride and Allenby (2003), and results in a total of $K+1=18$ parameters. The model is specified as follows

$$\begin{aligned} &V_{ijt} = e^{x_{ijt}^{'}\beta_i + \varepsilon_{ijt}} \quad \varepsilon_{ijt} \sim EV\left(0, \mu_i\right) \\ &\theta_i \sim N\left(\overline{\theta}, \Sigma_\theta\right) \ where \ \theta_i = \left[\beta_i^{'} \quad \mu_i^{*}\right]^{'} \end{aligned} \tag{7}$$

Where $t=1,\ldots,T$ is an index on the observations recorded for a given individual. The decision rule is based on (1) and the choice probabilities are based on (2). The sampler is a basic MNL sampler with a Metropolis-Hastings step to handle the non-conjugate mixture of the normal prior and the logit likelihood. Details on the MNL sampler can be found elsewhere in the literature, thus we do not elaborate them here (Allenby and Lenk 1994; Arora et al. 1998; Train 2003). We use the standard diffuse normal-inverted Wishart hyper-prior structure for $\overline{\theta}$ and Σ_θ. The prior mean of $\overline{\theta}$ is set to zero and the prior variance of $\overline{\theta}$ set to be arbitrarily large, in this case, $10^6 \times I_{(K+1)\times(K+1)}$. The prior degrees of freedom for Σ_θ is set to $\left(K+2\right)$ and the prior scale of Σ_θ is set to $I_{(K+1)\times(K+1)}$. In estimation, we exponentiate the draws of μ_i^* when computing the logit likelihood in the Metropolis-Hastings step. This implies the prior on μ_i is log-normal. Since this transformation is not a function of the location or scale of the normally distributed generating variable, no further adaptations to the sampler are required. We use the negative of ln(price) in the likelihood. We compare estimates of the distribution of θ_i from this parameterization with that of the linearized parameterization, where θ_i is replaced with $\Phi_i = \left[\varphi_i^{'} \quad \gamma_i^{*}\right]^{'}$. Here, the choice probabilities are based on (3). For consistency, we use a normal prior for φ_i, a log-normal prior for γ_i, and the negative of ln(price) in the likelihood. The same normal-inverted Wishart hyper-prior structure used for $\bar{\theta}$ and Σ_θ is used for $\bar{\Phi}$ and Σ_Φ. Estimates of the structural β_i's are computed as $\frac{\varphi_i}{\gamma_i}$ on each iteration of the sampler.

Study 2: Midsize Sedans

The second data set is CBC data on midsize sedans. The data were provided by a major automotive manufacturer. Respondents qualified for participation in the study on the basis of the vehicle they currently own, their intention to purchase a midsize sedan, their willingness to consider import manufacturers, and other socio-economic information. A total of 333 respondents participated in the study. Each respondent completed 15 choice tasks, with each task consisting of three sedans. The no-buy option was not included in this study.[4] There are no conditional pricing relationships in the experimental design. The average prices for the three alternatives across all the tasks is $22,000. Since the alternatives are all midsize sedans and there are no conditional pricing relationships, we have less reason to expect a non-linear response to price changes. Indeed, preliminary analyses reveal here that models using price fit the data better than those using log price. Therefore, these data are used to estimate the model based on surplus maximization.

We also have some demographic data describing the respondents. The available data include a binary indicator of gender that is one if the respondent is male, a binary

[4] Modeling the outside good via the no-buy option does eliminate specification error and allows the model to capture the market expansion effect. However, neither of these issues is of primary concern to the parameterization issue. Regardless of whether the outside good is modeled, linear utility models imply that WTP for an increase in an attribute level is estimated as the ratio of attribute coefficient to price coefficient, and the points regarding duality under different parameterizations remain valid.

education indicator that is one if the respondent has obtained a bachelor's degree or higher, a binary indicator that is one if the individual is married, the respondent's age, and the respondent's income. Age and income are measured via an interval scale. For age, the categories range from 1-9, with the lower bound for "under 20", the upper bound for "55 and over" and intervals of 5 years on the interior points. For income, categories range from 1-15, with the lower bound for "under $10,000", the upper bound for "$75,000 and over" and intervals of $5,000 on the interior points. The sample is 55% male, 71% have a bachelor's degree or higher, and 74% are married. The mean age score is 5.5 (std. deviation 2.1), which corresponds to a mean age of approximately 40. The mean income score is 8.21 (std. deviation 3.57), which corresponds to an approximate mean income in the neighborhood of $40,000-$45,000. The income distribution is slightly skewed towards the upper end of the distribution.

WTP is expressed as a linear function of the K non-price attributes. For identification, the make/model VW Passat is dropped, as are the lowest level of each of the safety features. To economize the number of parameters, we linearize the audio and engine attributes. This results in a total of $K+1=10$ parameters.[5] The model is specified as follows.

$$
\begin{aligned}
&C_{ijt} = x_{ijt}^{'}\beta_i + \varepsilon_{ijt} \quad \varepsilon_{ijt} \sim EV\left(0, \mu_i\right) \\
&\theta_i = \Delta_\theta z_i + \eta_i \quad \eta_i \sim N(0, \Sigma_\theta) \quad \text{where} \quad \theta_i = \left[\beta_i^{'} \quad \mu_i^{*}\right]^{'}
\end{aligned}
\tag{8}
$$

The decision rule is based on (4) and the choice probabilities are based on (5). The sampler is again a basic MNL sampler with a demographic regression step (Rossi et al.

[5] We acknowledge that it may be inappropriate to linearize the engine and audio attributes. However, our goal was to reduce the total number of parameters to a number smaller than the number of tasks per individual. This should reduce concerns about poor identification at the individual level that may be present in Study 1, which estimates 18 parameters on 12 choices per individual. Since the goal of this paper is to highlight issues involved in model parameterization, linearizing the attributes is of less concern.

1996). Again, since the details are readily available in the literature, we do not elaborate them here. We use a normal hyper-prior for Δ_θ and an inverted Wishart hyper-prior for Σ_θ. The prior mean of Δ_θ is set to zero and the prior variance of Δ_θ is set to $10^6 \times I_{(K+1)\times(K+1)}$. The prior degrees of freedom of Σ_θ is set to $(K+2)$ and the prior scale of Σ_θ is set to $I_{(K+1)\times(K+1)}$. As in (7), the prior on μ_i is log-normal and the negative of price is used in the likelihood. The vector z_i is a $p \times 1$ vector consisting of an intercept and $p-1$ demographic variables. In this SUR structure, the Δ matrix is a $(K+1)\times p$ matrix of regression coefficients.

We compare estimates of WTP and demographic regression coefficients based on the nonlinear surplus model with those based on the linear normalized utility model. As before, we replace $\frac{\beta_i}{\mu_i}$ with φ_i and $\frac{1}{\mu_i}$ with γ_i. Here, we specify the priors as

$$(9) \qquad \Phi_i = \Delta_\Phi z_i + \kappa_i \quad \kappa_i \sim N\left(0, \Sigma_\Phi\right) \text{ where } \Phi_i = \left[\varphi_i' \quad \gamma_i^*\right]'.$$

As before, the same normal-inverted Wishart hyper-prior structure used for Δ_θ and Σ_θ is used for Δ_Φ and Σ_Φ. Individual-level WTP estimates are calculated by computing $\beta_i = \frac{\varphi_i}{\gamma_i}$ at each iteration of the sampler. We can also compare the ability of demographic covariates to explain variation in the structural WTP estimates and variation in scale to their ability to explain variation in the normalized utility model parameters. It is well known and readily apparent from the re-parameterization of (5) that the utility model parameters commingle preference with scale (Swait and Louviere 1993). Furthermore, demographic covariates have been found to be of limited explanatory power in choice models based on linearized utility specifications (Rossi et al. 1996; Ainslie and Rossi 1998).

Results

In both studies, parameter estimates are calculated based on T-2 choice tasks. We also compute the in-sample fit via the harmonic mean estimator of the MLL (Newton and Raftery 1994). To assess out of sample performance, we use the last two tasks as holdout tasks. Using these, we compute the frequency of correct predictions over both holdout tasks and the probability of the chosen alternatives in each holdout task. These statistics are computed at each iteration of the sampler, and averaged over posterior point estimates. The hit probabilities are averaged over the holdout tasks, as well. In both studies, we use a burn-in of 20,000 iterations. Time series plots of model parameters indicate convergence. We ran the chains for an additional 10,000 iterations, keeping every 10th iteration for a total of 1,000 iterations for posterior inference.

Table 3 presents the posterior estimates of $\bar{\theta}$ and $\bar{\Phi}$ for the utility-per-dollar model.[6] Parameters with 95% coverage intervals that do not contain zero are listed in bold. The MLL's considerably favor the linear specification. In the tests of predictive validity, the holdout likelihoods, and hit probabilities slightly favor the linear specification while the hit frequencies favor the nonlinear specification. The differences across specifications, though, are very small. We cannot compare the magnitudes of $\bar{\theta}$ and $\bar{\Phi}$. However, there are a few differences in the precision of the estimates across the nonlinear and linear specifications.

[6] For brevity, we omit the estimates of the covariance matrices for both Study 1 and Study 2. Like the estimates of the means, the magnitudes cannot be compared. In both studies, the variance estimates of results indicate significant heterogeneity in the parameters of both specifications. Furthermore, in both studies, more of the nonlinear model parameters exhibit significant covariance. The results are available upon request.

---Insert Table 3 Here---

Table 4 compares the distribution of estimates of the structural parameters β_i and μ_i from the nonlinear and linear specifications of the utility-per-dollar model. We also compare the "price coefficient" across both specifications, $\frac{1}{\mu_i}$ for the nonlinear model and γ_i for the linear model. Comparing means and medians along with the ranges, Table 4 shows that the distribution of $\frac{\varphi_i}{\gamma_i}$ is generally more skewed and disperse compared to β_i. The correlations of the estimates across specifications, shown in the last column of the table, clearly indicate that the alternative parameterizations do not result in the same estimates of the structural parameters. The average correlation in the estimates is 0.74, with some correlations as low as 0.45.

---Insert Table 4 Here---

These results establish that linear re-parameterizations in conjunction with standard priors can lead to estimates of structural parameters that are skewed and prone to outliers. In the utility-per-dollar model, though, it is difficult to judge the face validity of the structural parameter, β. In this model, the partial derivative of V_j depends not only on β_k, but on the vectors x_j and β, too. It is entirely possible that the true distribution of the structural parameters is skewed. In the surplus model, though, β has precise economic meaning. This enables us to better judge the face validity of parameter estimates across model specifications.

Tables 5a-5c present fit statistics and estimates of the hierarchical regression coefficients Δ_θ and Δ_Φ. Parameters with 95% coverage intervals that do not contain zero are listed in bold. The findings on model fit mirror those of Study 1. In addition, demographic covariates appear to be of limited use in explaining variation in either the linear or the nonlinear specifications. However, it is interesting to note that demographics explain more of the variation in γ_i^* than μ_i^*.

---Insert Table 5 Here---

Table 6 compares the distribution of estimates of the structural parameter β_i from the nonlinear and linear specifications of the utility-per-dollar model. These are computed the same way as in Study 1. Table 6 also presents estimates of the coefficient on price from both specifications. As in Study 1, the means and medians along with the ranges show that the distribution of $\frac{\varphi_i}{\gamma_i}$ is more skewed and disperse compared to β_i. In this study, the means from the linear specification are consistently higher than the nonlinear specification. This is most likely due to the presence of outliers. For example, the linear specification 75th percentile scores for the Japanese make/models are between \$23,000 and \$33,000 (versus \$10,000-\$13,000 for the nonlinear model). The maximum values are between \$118,000 and \$169,000 (versus \$20,000 to \$25,000 for the nonlinear model). The implication of the linear model estimates is that for 25% of the respondents, the required compensation to induce indifference between similarly equipped Japanese sedans and the VW Passat is, at minimum, on the order of the sedan purchase prices, and at maximum, in the \$100,000's. In contrast, the maximum compensations for the nonlinear model is on the order of the purchase prices. These estimates are vastly different and would imply vastly different pricing policies. We argue that the estimates from the linear model lack face validity.

---Insert Table 6 Here---

Consider the other non-price features. We would expect most respondents to have positive value for increased engine performance, increased audio/navigation capability, and safety features. While some respondents may have negative value, indicating a preference for fuel economy and low emissions or an aversion to audio/navigation or safety technology, we'd expect these consumers to be in the minority. As Table 7 shows, the nonlinear specification implies between 5% and 10% of respondents have negative WTP's for these features, indicating they would require monetary compensation to be held utility equivalent given an increase in engine performance or audio/navigation capability, or the addition of safety features. The maximum compensations are in the thousands of dollars, with none in excess of $2,000. In contrast, the linear specification implies between 13% and 23% have negative WTP's, and the maximum compensations range from $9,620 to $21,620. Both the number of respondents requiring compensation and the magnitudes of the compensation implied by the linear model seem excessive.

---Insert Table 7 Here---

5. Monte Carlo Study

The findings from the two CBC studies exhibit the following consistencies. In both cases, the in-sample fits of the linear specification are considerably better than the nonlinear specification. In both cases, the hit frequencies favor the nonlinear specification. The hit probabilities are mixed; they are equal for the utility-per-dollar model and favor the linear specification of the surplus model. The estimates of the structural parameters from the linear specifications are more skewed than those of the nonlinear specifications and appear to lack face validity. Since we ultimately do not know the true value of the structural parameters, though, it is difficult to say with certainty that the linear model estimates are imprecise.

Using synthetic data sets, we wish to examine issues of fit and parameter recovery of the nonlinear and linear specifications more closely. For brevity, we will focus on the surplus maximization model. We use the design matrix from the CBC sedan data to generate synthetic data according to the decision rule and choice probabilities of surplus maximization. We generate two synthetic data sets. In the first, we use only the make/models and price, for a total of 5 covariates. In the second, we use the full design matrix for a total of 10 covariates. We calibrate the models on the first 13 choice occasions for each person and use the last two occasions for tests of predictive validity.

We generate the data sets using a value of 1 for the mean and variance of both μ_i^* and β_i. While this is an arbitrary choice, the value of 1 is close to the estimates for the mean and variance of μ_i^* in the CBC sedan data. We estimate both the nonlinear and linear models for each synthetic data set. The samplers are established in the same way as those in Section 4, with normal priors on $\left[\beta_i' \quad \mu_i^*\right]'$ and $\left[\varphi_i' \quad \gamma_i^*\right]'$, and the standard diffuse normal-inverted Wishart hyper-prior structure. We will examine both model fits and the efficacy of the nonlinear and linear specifications to recover $\left[\beta_i' \quad \mu_i^*\right]'$.

Table 8a presents the fit statistics of each parameterization across the two synthetic data sets. In both cases, the MLL's favor the linear model, despite the nonlinear model's consistency with the data generating process. The hit frequencies consistently favor the nonlinear model, and the hit probabilities are basically the same. These findings are largely consistent with the findings on model fit in Section 4.

---Insert Table 8a Here---

Table 8b presents, for each parameterization across the data sets, the mean absolute deviation (MAD) and the mean squared deviation (MSD) of the individual-level estimates of $\left[\beta_i^{'} \quad \mu_i^{*}\right]^{'}$. For the linear model, β_i is computed at each iteration of the sampler as $\frac{\varphi_i}{e^{\gamma_i^*}}$ and μ_i^* is computed as $\log\left(\frac{1}{e^{\gamma_i^*}}\right)$. The MAD's and MSD's are calculated at each iteration of the sampler and averaged over iterations. We then average the estimates over individuals. In both cases, the nonlinear model 95% coverage intervals contain the true values of the mean and variance of the individual-level parameters. In terms of recovery of the data generating values of $\left[\beta_i^{'} \quad \mu_i^{*}\right]^{'}$, the MAD's and the MSD's suggest the nonlinear model consistently recovers more precise estimates of β_i than the linear model. Note that the MSD's are, on average, 5-7 times larger for the linear model. Since the MSD's heavily penalize large deviations, this suggests the linear model is more prone to outlier estimates. Interestingly, in both data sets, the linear model seems to do a slightly better job of recovering the data generating values of μ_i^* than the nonlinear model, although the discrepancies are minor relative to those of the β_i parameters.

---Insert Table 8b Here---

Table 8c presents the variance estimates of the individual-level parameters. The linear model estimates drastically overstate the variance of the individual-level β_i parameters, and seems to understate the variance in μ_i^*. Indeed, for both cases, the linear model 95% coverage intervals for the variance of β_i and μ_i^* do not contain the data generating value of 1. In contrast, the coverage intervals for all the nonlinear model parameters contain the data generating values. The findings on MSD and the variance of the individual-level β_i parameters provide more evidence that the linear model estimates of β_i are over-dispersed and prone to outliers. This is a direct result of calculating the estimates as ratios of random variables.

---Insert Table 8c Here---

To summarize the Monte Carlo results, we find that the linear model has superior in-sample fits. Out of sample, the models seem to perform quite similarly; one may even give a slight edge to the nonlinear model based on hit frequencies. While fit statistics often play a large role in choosing from alternative choice model specifications, parameter recovery is equally important. In fact, in using CBC to design new products or test candidate designs, the model parameters are often the most important analytic output. In terms of parameter recovery, we have demonstrated that despite the disadvantage in in-sample fit, the nonlinear model has a decisive advantage in recovering the structural parameters of the data generating process. Similar discrepancies in model fit and parameter recovery have been noted elsewhere (Andrews et al. 2003; Liechty et al. 2003). The Monte Carlo findings are quite consistent with the findings from the two CBC studies presented in Section 4. Together, the findings suggest that researchers interested in structural parameters (WTP estimates are frequently a desired output from CBC studies, for example) should employ the nonlinear model rather than its re-parameterized linear counterpart.

6. Summary and Conclusion

We have demonstrated that when modeling choice in a setting with random coefficients, parameterization can have an impact on model fits and parameter estimates. This is in contrast to models with homogeneous coefficients or fixed effect approaches that leave both unaffected. Both Bayesian and classical approaches to random coefficient models requires the researcher choose a distribution for heterogeneous model parameters. In practice, Bayesian analysts typically utilize standard prior distributions, such as the multivariate normal, applied to parameters in a linear utility framework with either price or log price entering directly. These parameterizations are alternate means of specifying structural models of choice. The structural parameters can be recovered via transformations of the re-parameterized model coefficients. However, the use of normal priors for model parameters, including price, implies that the distribution of the structural parameters is a ratio of normals. In addition to possessing some rather undesirable properties, this can lead to invalid estimates of the structural parameters if the posterior distribution of the price coefficient contains non-negative values. While this can be remedied via a more theoretically appropriate choice of prior, the problem of skewed estimates that are prone to outliers with little face validity remains. There is also no straightforward way to assess the impact of strategic covariates on structural parameters estimates from the re-parameterized model. Any two-stage regression approach using the recovered structural parameters would have to confront the problem of outliers.

Using two CBC data sets, we have shown that linear parameterizations applied to heterogeneous choice models with standard priors results in very disperse estimates of the distribution of underlying structural parameters. These estimates appear to lack face validity. For example, the linear model applied to data on midsize sedan choice suggest that there are respondents with WTP values in the hundreds of thousands of dollars for the attributes they associate with some of the sedans in the study. Despite resulting in what appears to be more sensible estimates of the distribution of heterogeneity in structural parameters, the nonlinear specifications result in worse in-sample fits in both studies.

Using two Monte Carlo data sets, we illustrate that even when the data generating process follows the nonlinear model, the linear model has better in-sample fits. However, hit frequencies slightly favor the nonlinear model while hit probabilities suggest no difference in predictive validity. More importantly, the nonlinear model recovers more precise estimates of the structural parameters. The linear model also drastically overstates the amount of heterogeneity in the data generating process. The findings suggest that researchers interested in structural parameters should employ the nonlinear model rather than its re-parameterized linear counterpart. While this choice is arbitrary in homogenous or fixed effect models, it clearly has an impact in models with continuous distributions of heterogeneity. In terms of limitations and future research, an obvious limitation is that the results are based on survey data. While some authors have maintained that carefully designed CBC experiments can mimic real world markets, the evidence has been mixed and it would be beneficial to test the models on market data.

Table 1a: Utility per Dollar Maximization

	MLE Estimates					
	Non-linear Model		*Linear Model*		*Linear Model: Implied Parameters*	
	$\hat{\beta}$	t-stat	$\hat{\varphi}$	t-stat	$\hat{\beta}$	t-stat[a]
Low Body	1.41	4.33	0.82	2.90	1.41	4.08
Medium Body	3.67	11.19	2.13	6.37	3.67	13.98
High Body	2.77	6.47	1.61	4.13	2.77	6.70
Manual Change	0.72	1.97	0.42	3.24	0.72	3.18
Auto Change	0.33	2.85	0.19	3.97	0.33	4.41
Preset List	0.43	2.71	0.25	4.16	0.43	3.95
Custom List	0.86	5.09	0.50	7.98	0.86	6.38
Input 1	0.41	1.05	0.24	1.60	0.41	1.59
Input 2	1.00	5.34	0.58	8.27	1.00	6.14
Input 3	0.07	0.21	0.04	0.33	0.07	0.33
Ops Feedback	0.73	3.95	0.42	7.03	0.73	5.36
2x Zoom	1.53	6.95	0.89	13.93	1.53	7.71
4x Zoom	2.12	7.95	1.23	14.67	2.12	7.94
Large Viewfinder	0.12	0.75	0.07	1.11	0.12	1.09
LCD	0.32	1.90	0.19	1.71	0.32	1.72
Viewfinder	0.44	2.13	0.25	2.13	0.44	2.14
LCD & Viewfinder	0.60	3.16	0.35	2.90	0.60	2.89
	$\hat{\mu}$	t-stat	$\hat{\gamma}$	t-stat	$\hat{\mu}$	t-stat[a]
ln Price	1.72	10.74	-0.58	-9.10	1.72	9.10
LL	7373.02		7373.02			

a. via Taylor series approximation

Table 1b: Surplus Maximization

	MLE Estimates					
	Nonlinear Model		*Linear Model*		*Linear Model: Implied WTP*	
	$\hat{\beta}$	t-stat	$\hat{\varphi}$	t-stat	$\hat{\beta}$	t-stat[a]
Camry	3.80	16.64	0.63	15.55	3.80	13.81
Accord	0.84	4.46	0.14	3.23	0.84	3.21
Taurus	-4.60	-16.56	-0.77	-17.00	-4.60	-14.63
Maxima	1.91	9.70	0.32	8.04	1.91	7.80
Engine	1.17	14.37	0.20	15.48	1.17	13.49
Audio	0.75	6.25	0.12	4.14	0.75	4.09
Safety 1	1.00	10.18	0.17	7.96	1.00	7.70
Safety 2	0.75	7.70	0.13	6.05	0.75	5.95
Safety 3	0.65	6.72	0.11	4.66	0.65	4.63
	$\hat{\mu}$	t-stat	$\hat{\gamma}$	t-stat	$\hat{\mu}$	t-stat[a]
Price	5.99	30.05	-0.17	-25.00	5.99	25.00
LL	-4660.40		-4660.40			

b. via Taylor series approximation

Table 2: Attributes and Levels

Camera Data		Sedan Data	
Attribute	***Levels***	***Attribute***	***Levels***
Body Style	Low Medium High	*Make/Model*	Toyota Camry Honda Accord Ford Taurus VW Passat Nissan Maxima
*Mid-Roll Change**	None Manual Automatic	*Engine*	4 cylinder; 1.8 L; 150 HP 4 cylinder; 2.4 L; 160 HP 6 cylinder; 3.0 L; 155 HP 6 cylinder; 3.0 L; 222 HP
*Annotation**	None Pre-Set List Customized List Custom Input Method 1 Custom Input Method 2 Custom Input Method 3	*Audio and Navigation*	Standard Audio Premium Audio Premium Audio with Navigation
*Camera Operation Feedback**	No Yes	*Antilock Brakes*	No Yes
Zoom	None 2X 4X	*Side Door/Window Curtain Airbags*	No Yes
Viewfinder	Regular Large	*Vehicle Skid Control*	No Yes
Camera Settings Feedback	None LCD Viewfinder LCD & Viewfinder		
Price (nested within camera type)	from $41 to $499	*Price ($1,000)*	$17.4 $18.9 $20.4 $21.9 $23.4 $24.9 $26.4

*Only available on APS camera

Table 3: Parameter Estimates and Fit Statistics: Utility-per-Dollar Model[1]

			Nonlinear Model	Linear Model
Attribute	*Level*	k	$\bar{\beta}$	$\bar{\varphi}$
Body Style	Low	1	**-5.78**	**-1.13**
	Medium	2	**4.29**	**2.34**
	High	3	**2.24**	**1.47**
Mid-Roll Change*	Manual	4	0.88	**1.08**
	Automatic	5	0.34	0.17
Annotation	Pre-Set List	6	**1.00**	**0.34**
	Customized List	7	**1.88**	**0.74**
	Custom Input Method 1	8	**-1.92**	**-1.32**
	Custom Input Method 2	9	**2.13**	**0.86**
	Custom Input Method 3	10	**-2.28**	**-1.44**
Camera Operation Feedback	Operation Feedback	11	**0.80**	**0.42**
Zoom	2X Zoom	12	**3.44**	**1.40**
	4X Zoom	13	**4.26**	**1.88**
Viewfinder	Large Viewfinder	14	-0.39	-0.13
Camera Settings Feedback	LCD	15	-0.17	**0.93**
	Viewfinder	16	-0.18	**0.94**
	LCD & Viewfinder	17	0.23	**1.14**
			$\bar{\mu}^*$	$\bar{\gamma}^*$
		18	**0.83**	**-0.51**
Fit Statistics				
MLL			-3930.3	-3740.2
Hit Frequency			272	267
Hit Probability			0.38	0.38

1. Parameters whose 95% coverage interval does not contain zero listed in bold.

Table 4: Distribution of Individual-Level Model Parameters and Correlations

		Nonlinear Model			Linear Model			
			β_i			$\frac{\varphi_i}{\gamma_i}$		
Attribute	*Level*	mean	median	range	mean	median	range	ρ
Body Style	Low	-5.77	-6.23	34.03	-1.71	-2.16	44.89	0.86
	Medium	4.29	4.58	17.70	5.00	4.39	20.56	0.88
	High	2.25	2.61	12.89	3.18	2.44	14.57	0.72
Mid-Roll Change	Manual	0.87	0.92	9.75	2.23	1.65	18.30	0.72
	Automatic	0.33	0.26	11.81	0.26	0.07	21.83	0.91
Annotation	Pre-Set List	1.00	1.07	3.55	0.63	0.49	7.77	0.73
	Customized List	1.88	1.95	4.98	1.47	1.33	8.65	0.77
	Custom Input Method 1	-1.93	-2.07	23.27	-3.19	-2.37	41.19	0.87
	Custom Input Method 2	2.13	2.22	8.98	1.71	1.32	12.45	0.86
	Custom Input Method 3	-2.28	-2.39	19.33	-2.98	-2.62	29.57	0.87
Camera Operation Feedback	Operation Feedback	0.80	0.66	6.50	0.70	0.42	12.75	0.80
Zoom	2X Zoom	3.44	3.56	9.74	2.53	2.32	12.20	0.72
	4X Zoom	4.26	4.25	11.61	3.19	2.97	19.58	0.71
Viewfinder	Large Viewfinder	-0.39	-0.50	6.79	-0.37	-0.28	11.15	0.80
Camera Settings Feedback	LCD	-0.18	-0.08	5.32	1.79	1.57	10.73	0.45
	Viewfinder	-0.19	-0.15	6.61	1.83	1.57	8.47	0.45
	LCD & Viewfinder	0.23	0.32	6.22	-1.71	-2.16	44.89	0.56
			$\frac{1}{\mu_i}$			γ_i		
Price	-ln(price)	0.45	0.37	1.37	0.75	0.57	2.51	0.67

Table 5a: Model Fit Statistics

Fit Statistics	*Nonlinear Model*	*Linear Model*
MLL	-2508.3	-2147.8
Hit Frequency	410	409
Hit Probability	0.56	0.58

Table 5b: Nonlinear Surplus Model---Posterior Means (Standard Deviations), Δ_θ

		Constant	*Male*	*College Educated*	*Married*	*Age*	*Income*	ρ^2
Ford Taurus	β_{1i}	**-3.73**	**3.22**	-2.10	3.00	**0.71**	0.17	0.07
Toyota Camry	β_{2i}	**7.17**	-1.24	1.55	2.13	**0.99**	-0.37	0.06
Nissan Maxima	β_{3i}	**5.13**	0.50	0.53	-0.43	**0.73**	0.08	0.02
Honda Accord	β_{4i}	**3.99**	-1.53	2.00	**2.69**	**0.68**	-0.28	0.05
Engine	β_{5i}	**1.85**	0.40	-0.50	-0.05	-0.03	**0.12**	0.07
Audio	β_{6i}	**0.81**	-0.28	-0.43	-0.14	0.01	0.01	0.07
Safety 1	β_{7i}	**2.25**	-1.13	-0.64	-0.02	0.00	0.02	0.12
Safety 2	β_{8i}	**1.50**	-0.94	0.51	0.02	-0.09	-0.01	0.14
Safety 3	β_{9i}	**1.37**	0.20	-0.30	-0.77	-0.09	0.09	0.10
	μ_i^*	**1.21**	-0.49	-0.19	-0.22	0.05	0.00	0.10

Table 5c: Linear Surplus Model---Posterior Means (Standard Deviations), Δ_Φ

		Constant	*Male*	*College Educated*	*Married*	*Age*	*Income*	ρ^2
Ford Taurus	φ_{1i}	**-1.14**	0.72	-0.73	1.13	**0.29**	0.03	0.05
Toyota Camry	φ_{2i}	**2.47**	0.23	0.74	**1.27**	**0.27**	**-0.14**	0.05
Nissan Maxima	φ_{3i}	**1.82**	0.60	0.31	0.35	**0.20**	0.00	0.02
Honda Accord	φ_{4i}	**1.52**	**-0.25**	0.71	**1.10**	**0.21**	-0.10	0.03
Engine	φ_{5i}	**0.65**	0.39	-0.17	0.04	-0.04	**0.04**	0.06
Audio	φ_{6i}	**0.29**	0.02	-0.13	0.01	-0.01	-0.01	0.00
Safety 1	φ_{7i}	**0.72**	-0.15	-0.18	0.20	0.02	-0.01	0.01
Safety 2	φ_{8i}	**0.46**	0.06	0.20	-0.06	-0.05	-0.01	0.01
Safety 3	φ_{9i}	**0.46**	0.18	-0.12	-0.34	-0.04	0.04	0.02
	γ_i^*	**-1.43**	**0.76**	**0.40**	**0.34**	**-0.08**	-0.04	0.16

Table 6: Distribution of β_i and $\frac{\varphi_i}{\gamma_i}$

		Nonlinear Model			*Linear Model*			
			β_i			$\frac{\varphi_i}{\gamma_i}$		
Attribute	*Level*	mean	median	range	mean	median	range	ρ
Make/Model	Ford Taurus	-3.72	-3.65	55.89	-7.34	-2.54	266.09	0.83
	Toyota Camry	7.20	7.57	40.21	19.76	11.27	288.49	0.78
	Honda Accord	5.14	6.05	40.63	14.46	7.31	237.67	0.79
	Nissan Maxima	4.00	4.48	37.37	10.45	5.79	255.23	0.78
Engine	Engine	1.84	1.82	6.37	4.12	2.54	51.89	0.69
Audio and Navigation	Audio/Nav	0.82	0.82	4.05	2.08	0.97	36.89	0.66
Antilock Brakes	Yes	2.25	2.30	8.13	5.95	3.69	72.40	0.66
Side Door/Window Curtain Airbags	Yes	1.51	1.48	6.28	3.30	1.62	49.47	0.58
Vehicle Skid Control	Yes	1.38	1.37	5.07	3.14	1.74	64.70	0.66
			$\frac{1}{\mu_i}$			γ_i		
Price	-Price	0.42	0.38	1.29	0.42	0.27	2.38	0.63

Table 7: Percent of Distribution with $\beta_i < 0$ and Minimum WTP: Sedan Features

Specification	*Nonlinear Model*		*Linear Model*	
	%<0	Min WTP	%<0	Min WTP
Engine	5%	-1.58	13%	-16.04
Audio and Navigation	10%	-1.14	23%	-9.62
Antilock Brakes	6%	-1.90	12%	-18.70
Side Door/Window Curtain Airbags	7%	-1.95	17%	-12.71
Vehicle Skid Control	8%	-1.19	18%	-21.62

Table 8a: Monte Carlo Model Fit Statistics

Covariates	5		10	
	Nonlinear	Linear	Nonlinear	Linear
MLL	-3269.00	-3261.70	-3080.90	-3069.40
Hit Frequency	411	399	405	387
Hit Probability	0.53	0.53	0.53	0.53

Table 8b: Monte Carlo Parameter Estimates: MAD and MSD

Covariates	5		5		10		10	
	MAD		MSD		MAD		MSD	
	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
β_1	0.95	1.93	1.44	11.49	1.15	2.18	2.11	11.63
β_2	1.02	2.03	1.63	12.10	1.31	2.51	2.73	15.75
β_3	1.07	1.95	1.81	11.13	1.02	2.09	1.66	10.86
β_4	1.18	2.07	2.22	11.68	1.04	2.08	1.71	11.08
β_5					0.84	1.34	1.16	4.38
β_6					0.94	1.63	1.43	6.43
β_7					1.03	1.98	1.65	9.66
β_8					1.00	1.98	1.57	9.42
β_9					1.05	1.88	1.75	8.28
μ^*	0.61	0.60	0.62	0.60	0.73	0.72	0.87	0.84

Table 8c: Monte Carlo Parameter Estimates: Variance of β_i

Covariates	5		10	
	Nonlinear	Linear	Nonlinear	Linear
β_1	0.70	10.70	1.18	10.80
β_2	0.74	11.24	2.06	15.39
β_3	0.99	10.63	0.83	9.67
β_4	1.48	11.08	0.89	10.47
β_5			0.93	4.92
β_6			1.05	6.47
β_7			0.87	9.09
β_8			0.77	9.13
β_9			0.94	8.04
μ^*	0.87	0.68	1.22	0.46

References

Ainslie, A. and P. Rossi (1998), “Similarities in Choice Behavior Across Product Categories,” *Marketing Science*, 17, 91-106.

Allenby, G. and P. Lenk (1994), “Modeling Household Purchase Behavior with Logistic Normal Regression,”*Journal of the American Statistical Association*, 89, 1218-1231.

Arora, Neeraj, Greg M. Allenby and James L. Ginter (1998) “A Hierarchical Bayes Model ofPrimary and Secondary Demand,” *Marketing Science*, 17, 29-44.

Andrews, R. A. Ainslie and I. Currim (2002), “An Empirical Comparison of Logit Choice Models with Discrete vs. Continuous Representations of Heterogeneity,” *Journal of Marketing Research*, 39, 479-487.

Cameron, T. and M. James (1987), “Efficient Estimation Methods for 'Closed Ended' Contingent Valuation Surveys,” *Review of Economics and Statistics*, 69, 269-276.

Gilbride, T. and G. Allenby (2003) “A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules,” *Marketing Science,* forthcoming.

Hanemann, M. and B. Kanninen (1998), “Statistical Considerations in Contingent Valuation” in *Valuing Environmental Preferences - Theory and Practice of the*

Contingent Valuation Method in the US, EU , and Developing Countries, I. Bateman and G. Willis, eds. Oxford University Press.

Liechty, J., D. Fong and W. DeSarbo (2003) "Dynamic Models Incorporating Individual Heterogeneity: Utility Evolution in Conjoint Analysis," working paper, Pennsylvania State University.

Jedidi, K. and J. Zhang (2002), "Augmenting Conjoint Analysis to Estimate Consumer Reservation Prices," *Management Science*, 48, 1350-1368.

---, S. Jagpal and P. Manchanda (2003), "Measuring Heterogeneous Reservation Prices for Product Bundles," *Marketing Science*, 22, 107-130.

Mehta, N., S. Rajiv and K. Srinivasan (2003), "A Structural Estimable Model of Forgetting in Memory Based Choice," working paper, University of Toronto.

Newton, M.A. and Raftery, A.E. (1994), "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, Series (B)* 56, 43-48.

Rossi, P., R. McCulloch, and G. Allenby (1996), "On the Value of Household Purchase History Information in Target Marketing," *Marketing Science*, 2, 321-340.

Swait, J. and J. Louviere (1993), “The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models,” *Journal of Marketing Research*, 30, 305-314.

Train, K. (2003) *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press.

Zehna, P.W. (1966) “Invariance of Maximum Likelihood Estimators”, *Annals of Mathematical Statistics*, 37, 744