Food Marketing Policy Center

A Structured Covariance Probit Demand Model

by Michael A. Cohen

Research Report Series

http://www.fmpc.uconn.edu

University of Connecticut
Department of Agricultural and Resource Economics

A Structured Covariance Probit Demand Model [†]

Michael A. Cohen

University of Connecticut

Food Marketing Policy Center

`michael.cohen@uconn.edu`

January 22, 2010

Abstract

This paper introduces a heterogeneous agent discrete choice probit demand model with a structural interpretation of product choice covariance designed to overcome two hurdles in discrete choice demand modeling. One hurdle is the curse of dimensionality implicit in covariance probit demand models and the other hurdle is the independence of irrelevant alternatives (IIA) implicit in logit demand models. The structured covariance probit exploits the fact that choice models rely on utility differences to achieve identification. The utility difference structure implied by the additive random utility model is imposed on the covariance matrix and requires just one parameter in addition to those specified in the deterministic component of consumer utility. As an additional advantage the structured covariance probit is a better out-of-sample predictor because it allows covariance to change according to characteristics of the market. To estimate the model the paper develops a Bayesian estimation approach. The model also incorporates a Dirichlet process prior over normally distributed consumer segment clusters to flexibly model demand heterogeneity. The new model is evaluated relative to the widely used heterogeneous consumer logit demand model. Sampling experiments confirm that the model performs well under misspecification. An empirical analysis demonstrates that the new probit model captures realistic unrestricted switching behavior whereas its logit counterpart exhibits restrictiveness inconsistent with the utility theory on which the model is based.

---

1 Introduction

An important aspect of differentiated product demand analysis is to employ a model that is adequately flexible yet easy to identify and estimate with typically available data. Research such as McFadden (1981) and Hausman and Wise (1978) developed discrete choice approaches, that adopt the random utility model (RUM). Other work applied these models to construct a demand system (Allenby, 1989; Berry, Levinsohn, & Pakes, 1995). Altogether, a probability choice model, such as the logit or probit, is specified at the consumer level and heterogeneity is introduced as a distribution over RUM parameters then aggregated to build a demand system. These models specify utility as a function of product characteristics, a practice that overcomes the curse of dimensionality inherent in product space demand systems.

Researchers widely adopted the logit because is admits closed form probabilities and the introduction of heterogeneity relaxes the well known independence of irrelevant alternatives (IIA) property at the market level.[1] The most unsavory fact is that the IIA problem still dictates consumer level behavior. The extent to which heterogeneity ameliorates IIA at the aggregate level is an empirical question and models that suffer from IIA at the consumer level should be used with caution and only after they have been determined to be robust in the context of the specific empirical setting to which they are applied.

On the other hand, the probit relaxes IIA by modeling choice covariance based on the relative location of product choices in product characteristic space. Probit probabilities require simulation of integrals over the multivariate normal density, however computational methods have advanced to overcome this challenge. However, if one estimates a full covariance matrix the number of elements to be estimated are directly proportional to the number of products in the market, consequently the curse of dimensionality rears its head.

Identification of any choice model relies exclusively on differences between consumer utilities (Train, 2003, p.11). This paper develops a new flexible probit demand specification that exploits the identification strategy further by introducing a structural interpretation of choice covariance to the probit model. The structural model dictates that elements of the covariance are a function of the differences between deterministic components of the random utility specification for each product. This approach requires the estimation of only one additional scaling parameter regardless of the number of products in the market.

This paper also constructs a hybrid sampling technique for efficient estimation of the model. The model and the Markov-Chain-Monte-Carlo (MCMC) simulation approach is tested in a sampling experiment to verify their performance under proper specification and misspecification. I supply further validation of the new probit with an empirical example that applies the model to a panel data set on household purchases

---

[1] The IIA restriction of the logit model has been well understood since Debreu (1960)

of lemon-lime sodas in the New York City Designated Marketing Area (DMA).

One component of this research is to compare the performance of the heterogeneous agent structured covariance probit demand model to the widely used heterogeneous agent logit demand model. In this vein I evaluate a model's ability to do so in a theoretically justifiable way. For example I demonstrate that IIA still presents itself after aggregation of the logit to the market level by conducting counterfactual experiments that compare logit elasticity estimates to those from the new probit demand model. This type of experimentation improves our understanding about the way discrete choice demand systems behave. My results are important to applied researchers because they demonstrate the appropriateness of logit or probit demand models for the various types of empirical work they conduct. It stands to reason that the extent to which IIA predetermines empirical results is extremely important to applied researchers using structural demand models to analyze differentiated product markets. This research also provides insight on how one can test for IIA in their own empirical analysis.

This is not the first work to place structure on a covariance probit, Yai, Iwakura, and Morichi (1997) uses geographic distance to model correlation for transportation choices. Recent work by Lenk, Dotson, Brazell, Otter, Maceachern, and Allenby (2009) proposes an ad hoc structure for probit covariance to capture the the similarity effect (Tversky, 1972, 2003), and use three different distance metrics to capture product similarity. Their conjoint study models covariance as a function of product similarity to estimate source of volume calculations when new products are introduced.

This paper extends the models of Yai et al. (1997) and Lenk et al. (2009) in two ways. First, it introduces the structured covariance probit to demand analysis. Second it specifies a Dirichlet process prior on the consumer heterogeneity distribution offering more flexibility to the demand model.

The entire literature that develops and applies discrete choice demand systems is vast and not all directly pertinent to the current discussion so I refer the interested reader to Ackerberg, Benkard, Berry, and Pakes (2007) for thorough coverage of its evolution. The remainder of the introduction provides specific accounts of the research lines that develop the probit, consider aggregation of choice models, and specify models of consumer heterogeneity.

Hausman and Wise (1978) are largely credited with the introduction of the conditional (or multinomial) probit model to economic analysis. Restrictions on Σ are required to ensure identification of the model. Bunch (1991) suggests one way to identify the model, showing that if the utilities in a choice set are differenced, all but one of the parameters of the Σ are identified. Bayesian model estimation is considered by McColloch and Rossi (1994) who propose a Gibbs sampler that navigates the unidentified parameter space, relying on the fact that post estimation, parameters are identified relative to one of the choices. Their ap-

proach normalizes the unidentified draws, post process, with the posterior variance of one of the parameters. McColloch, Polson, and Rossi (2000) suggest a Gibbs sampler to navigate the identified parameter space, but find this identified sampler performs slower and less efficiently than the McColloch and Rossi (1994) method. These approaches rely on the data augmentation techniques of Albert and Chib (1993) who develop a method to circumvent evaluation of the probit likelihood. To accomplish this they show that one can draw the latent utility variable from a multivariate normal distribution conditional of the data and parameters and subsequently use the draws as dependent variables in a regression to estimate utility model parameters.

The aggregation properties of these market models are examined by Allenby and Rossi (1991). They establish three conditions under which the use of aggregate logit models is theoretically justifiable. The first is that all consumers are exposed to the same marketing mix variables. The second is that the brands under consideration are close substitutes. And the third, is that the distribution of prices is not concentrated at an extreme value. Their work also documents that when aggregating up to the market level the extent to which IIA is preserved depends on the distribution of consumer tastes. They show in some cases that the logit substitution patterns are preserved though aggregation.

Ackerberg et al. (2007) explain that obtaining aggregated demand from a distribution of household preferences has two important advantages. One, it allows the researcher to use the same framework to study demand in different markets, and two, heterogeneous-agent-based demand systems are readily able to analyze the distributional impacts of policies proposed to effect goods marketed. While the representative agent approach might generate reasonable demand estimates in one market, the advantage of aggregation from a distribution of consumer preferences allows the accommodation of often large differences in the distribution of demographic characteristics in one market compared to another. This avoids having an estimated model that fits well in one market but performs poorly in another.

Appreciating the varied tastes of consumers in the demand model is not only a more realistic representation of market demand it also improves the theoretical properties of the logit demand model. Some researchers have observed that incorporating a flexible specification of consumer utility in the market level logit model relaxes IIA at that level. Chintagunta, Jain, and Vilcassim (1991) use consumer-level panel data to estimate a model that interacts demographic characteristics of the consumers. This effectively makes the distribution of consumer preferences depend on the products they did not purchase by introducing a nonlinear component. Well known research by Berry et al. (1995) introduces a demand model that allows consumers to value choice characteristics differently by modeling a distribution of persistent taste shocks across the population of consumers. They effectively exploit the same interactions as Chintagunta et al. (1991), however they apply the contract mapping suggestion of Berry (1994) to deal with the nonlinear id-

iosyncratic part of the model and exploit the unobserved product characteristic to estimate the model with aggregate data. McFadden and Train (2000) show that these mixed logit models are highly flexible and in theory can approximate any random utility model.

Bayesian semi-parametric approaches accommodate estimation of flexible density specification. The Dirichlet process prior mixes normal densities to achieve multi-modality, asymmetries, and fattened tails. By mixing enough normal distributions one can build any shape distribution just as if one were piling mass on a flat surface. Tchumtchoua (2008) models the distribution of consumer preferences using a Dirichlet process which vary with consumer demographic characteristics. Burda, Harding, and Hausman (2008) presents a Bayesian Markov-Chain-Monte-Carlo method for estimating mixed the logit and mixed probit. The Dirichlet process prior allows them to estimate a joint density of the model preference coefficients to reduce the parametric restriction on the heterogeneity. Dube, Hitsch, and Rossi (2009) specify a Dirichlet process prior for heterogeneity to rule out an alternative explanation for inertia in consumer purchase decisions. Specifying a flexible distribution of heterogeneity eliminates critiques that cite inflexible specification of the consumer preference distribution as the reason for spurious results testifying to state dependence in consumer utility functions. For example if the distribution of consumer preference is flexible consumers that purchase Mountain Dew all the time do so because they have a taste for caffeine. If this is the case then consumers are not purchasing the product just because they did last shopping trip (inertia).

This remainder of this paper is structured as follows. Section 2 introduces the model and the demand response structure it implies. Section 3 states the procedures for Bayesian model inference using Markov Chain Monte Carlo simulation methods. Section 4 conducts a sampling experiment. Section 5 provides and empirical example that demonstrates differences between the new model and its popular logit counterpart. Section 6 concludes.

2   Model

This section introduces the new demand model. It begins by deriving the model from an underlying theory of consumer utility closely following the exposition of Berry et al. (1995). Next it states the structural representation. Then it derives the demand elasticities.

2.1   Utility and Demand Model

Aggregating a discrete choice model of individual behavior over a population gives one a demand system. The utility an individual derives from a specific product is a function of individual characteristics, ζ, and a

set of observed and unobserved product characteristic, respectively, $[x, \xi]$. The utility derived by consumer, i, for product, j, is:

$$U(\zeta_i, x_j, \xi_j; \theta), \tag{1}$$

θ parameterizes the utility model, including parameters that characterize the distribution of consumer characteristics, and parameters that characterize the utility function, conditional on consumer characteristics. Given preferences the consumer chooses the product that offers the highest utility.

$$U(\zeta_i, x_j, \xi_j; \theta) \geq U(\zeta_i, x_k, \xi_k; \theta), \text{ for } k = 0, 1, \dots, J, \tag{2}$$

where k indexes other products under consideration for purchase and 0 represents the no purchase outside option. Including the outside option is essential for specifying aggregate demand as a function of observed product characteristics. The set,

$$B_j = \{\zeta : U(\zeta, x_j, \xi_j; \theta) \geq U(\zeta, x_k, \xi_k; \theta), \text{for } k = 0, 1, \dots, J, \tag{3}$$

characterizes the set of values for ζ that motivates choice of good j. Assuming that ties occur with zero probability, the market share for good j is:

$$S_j(x_j, \xi_j; \theta) = \int_{\zeta \in B_j} dP(\zeta). \tag{4}$$

$P(\zeta)$ is the probability density function of consumer characteristics. Grouping all s_j into a J-dimensional vector, $s(\cdot)$, demand for each product is given by, $Ms(x_j, \xi_j; \theta)$, where M is the size of the market.

Under the practice that utility is additively specified, commonly referred to a the additive random utility model (ARUM), it may be expressed as:

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \text{ for } j = 0, 1, \dots, J. \tag{5}$$

Here V_{ij} is the deterministic component of utility and ε_{ij} denotes the random component with $\mathbb{E}[\varepsilon_{ij}] = 0$. The deterministic component is typically specified as:

$$V_{ij} = x_j' \beta_i + \xi_j. \tag{6}$$

ξ_j is the mean unobserved component of utility for product j, and heterogeneity is introduced through

β_i. The β_is are consumer i's marginal utility for product j's observed characteristics, x_j, conditional on consumer characteristics, ζ. An affine transformation induces the mapping of ζ into β_is,

$$\beta_i = \bar{\beta} + z_i'\delta + \nu_i. \tag{7}$$

The marginal utility parameter mean is $\bar{\beta}$, and the deviation from the mean is generated by, observed demographic characteristics, z_i, translating parameters δ, and unobserved characteristics, ν_i. The unobserved shock, ν_i, has a specified density that reflects the underlying distribution of the β_is, namely $P(\zeta)$.

Equation 2 implies that,

$$U(\zeta_i, x_k, \xi_k; \theta) - U(\zeta_i, x_j, \xi_j; \theta) \leq 0, \text{ for } k \neq j. \tag{8}$$

Equations 6 and 7 together along with a probability model of consumer choice yields:

$$Pr[j] = Pr[\varepsilon_{ik} - \varepsilon_{ij} \leq V_{ij} - V_{ik} \text{ for } k \neq j] \tag{9}$$
$$= Pr[\tilde{\varepsilon}_{ikj} \leq V_{ij} - V_{ik} \text{ for } k \neq j]. \tag{10}$$

This equation clearly demonstrates that only the difference in utilities matters in identifying the choice model. This underscores the fact that correlation in choice is determined by differences in utility, and more precisely differences in the deterministic component of utility. Assuming $\tilde{\varepsilon}_{ikj}$ comes from a probability distribution F, market shares are given by:

$$S_j = \int_{\zeta \in B_j} \int I(\tilde{\varepsilon}_{ikj} \leq V_{ij} - V_{ik} \forall k \neq j) dF(\tilde{\varepsilon}_{ikj}) dP(\zeta). \tag{11}$$

If F is assumed to be the multivariate normal density then the underlying choice model is the probit, and $\tilde{\varepsilon} \sim \mathcal{N}[0, \Sigma]$. Exploiting the fact that the correlation in choice must be determined by differences in utility, elements of the covariance matrix Σ are modeled as a function of differences in the deterministic component of utility, $V_{ij} - V_{ik}$. Not only does this practice follow from the utility theory that underlies the model it also ensures that the full covariance matrix is identified.

2.2 Structural Model for Choice Covariance

A utility difference can be thought of as a distance in utility space. That is, after controlling for the observable product characteristics one expects choices to have a higher degree of correlation for products that

the consumer considers as closer alternatives. An appealing idea is to model error correlation as a function of the distance between the utilities a consumer gets from each product. Lenk et al. (2009) apply this idea to a conjoint model and argue effectively this practice captures similarity effects (spacial differentiation).

The off-diagonal elements of Σ are modeled for individual i and choice set c, they take the form:

$$\Sigma_{ic} = \begin{bmatrix} 1 & \cdots & \sigma_{1j}^{(i)} \\ \vdots & \ddots & \vdots \\ \sigma_{k1}^{(i)} & \cdots & 1 \end{bmatrix} \tag{12}$$

The off-diagonal elements are the covariance between choice alternatives k and j. notice that the diagonal elements are 1. This is because the exponential function is applied to utility differences to ensure symmetry and positive definiteness of the matrix, and since the distance between a product and itself is 0 we know $exp(0) = 1$. The specific structure I impose on elements of the covariance matrix is:

$$\sigma_{kj}^{(i)} = exp[\frac{-d_{kj}}{exp(\theta_i)}]. \tag{13}$$

d_{kj} is the difference in deterministic utility for choice alternatives k and j. θ_i scales d_{kj}. On one hand a consumer with a large θ_i views relative product utility as an important determinant of switching. A consumer with a θ_i close to zero views products independent of their characteristics, just like the logit. θ_i is jointly distributed with the β_is under the density, $P(\zeta)$, and mapped from ζ in analogous fashion. The difference between deterministic utilities of a choice alternatives is,

$$d_{ikj} =| (x_k - x_j)'\beta_{\mathbf{i}} | . \tag{14}$$

A variation on this difference is considered by Lenk et al. (2009),

$$d_{kj} = \sum_{m=1}^{M} | (x_m^k - x_m^j)\beta_m | . \tag{15}$$

This metric precisely measures the sensitivity of the consumer to product characteristics. In equation 15 utility difference is viewed as an index weighting the distance between characteristics by their marginal utilities. This metric contrasts the first metric because is views characteristics independently rather than as a whole, and potentially generates different covariance estimates. Going forward I refer to the two distance metrics as the D3 and D2 metrics respectively.

2.3   Marginal Effects

As one expects, a model that captures flexible substitution patterns improves elasticity model flexibility. Better elasticity estimates mean more accurate analysis of a differentiated product market and therefore more reliable measures of consumer response. Because the probit captures similarity effects at the consumer level, an additional dimension over the random coefficients logit, the level of substitution pattern richness relaxes any concern for IIA.

In random coefficient logit demand systems similarity effects are captured by aggregation over the distribution of heterogeneity. The intuition is that consumers with similar sensitivity to a product characteristic, such as price, have similar switching behavior. Elasticity estimates calculated by integrating over the distribution of consumer sensitivity parameters produce "flexible" patterns of substitution. This basis allows the model to capture similarity effects at the market level if one appreciates that the aggregate effect has consumers buying Coke moving to Dr. Pepper because someone with a similar taste profile also chooses it. This means that we are lining up consumer's individual responses in a way dictated by the distribution of sensitivities to product characteristics.

Under the logit assumption a consumer switches independent of product characteristics. Recall that the IIA property of the model dictates switching behavior to be independent of product characteristics and dependent only of the relative magnitudes of choice probabilities, consequently there is no similarity effect. At the market level elasticities are constrained by the specification for the distribution of heterogeneity. This point begs whether adequate information is contained in the data about consumer characteristics to strongly tie consumers to the distribution. If not it stands to reason that estimates of the heterogeneity distribution and therefore estimates of market switching behavior are not accurate. This remains an empirical question.

The consumer response can be decomposed by evoking the chain-rule for differentiation. The single response results from a change in share which is brought about by the change in utility, and the change in utility brought about by a change in a product characteristic. In the following utility, u, is normalized:

$$\frac{\partial s_{ij}}{\partial x_j} = \frac{\partial s_{ij}}{\partial u_{ij}} \frac{\partial u_{ij}}{\partial x_j}. \tag{16}$$

For the linear RUM $\frac{\partial u_{ij}}{\partial x_j} = \beta_i$. The probit consumer share response to a change in utility is:

$$\frac{\partial s_{ij}}{\partial u_{ik}^{(j)}} = \phi(u_{ik}^{(j)}) \times \Phi[\cdots, \frac{u_{ik}^{(j)} - r_{lk}^{(j)} u_{il}^{(j)}}{(1 - [r_{lk}^j]^2)^{1/2}}, \cdots ; R_{.k}^{(j)}], \quad k \neq l \tag{17}$$

where $\phi(\cdot)$ is the the univariate standard normal density. $\Phi(\cdot; R_{.k}^{(j)})$ is a mean 0 multivariate normal distri-

bution function with covariance matrix, $R_{.k}^{(j)}$, equal to a differenced probit covariance Σ with out the row and column corresponding to good k, the r_{lk}'s are the elements of the removed vectors. The derivation of this expression is included in appendix A of this paper. These derivatives can then be used to calculate own and cross response, they appear respectively:

$$\frac{\partial s_{ij}}{\partial x_j} = \beta_i \sum_j \frac{\partial s_{ij}}{\partial u_{ik}^{(j)}} [\delta_{jj}^{(j)}]^{-1/2}. \tag{18}$$

$$\frac{\partial s_{ij}}{\partial x_k} = -\beta_i \frac{\partial s_{ij}}{\partial u_{ik}^{(j)}} [\delta_{jj}^{(j)}]^{-1/2}.$$

Where the $\delta_{jj}^{(j)}$s are the diagonal variance elements of the probit covariance matrix. Recognize that $\frac{\partial s_{ij}}{\partial u_{ik}^{(j)}}$ is the product of a probability distribution function and a cumulative distribution function therefore it is positive. Further recognize that $\delta_{jj}^{(j)}$ is a variance and therefore positive. Hence the sign of the derivative is determined solely by the sign of the consumer sensitivity to x_j, namely β_i. These expressions are used in the subsequent section to derive market elasticities for the probit demand model.

Market demand elasticity is calculated by integrating up the own-product characteristic responses of each consumer over the distribution of preferences in the market population:

$$E_{jj} = \frac{S_j}{x_j} \int_i \frac{\partial s_{ij}}{\partial x_j} dP(\zeta). \tag{19}$$

Similarly the cross-product-characteristic market demand elasticity is calculated by integrating up the cross-product-characteristic responses of each consumer over the distribution of preferences in the market population:

$$E_{jk} = \frac{S_j}{x_k} \int_i \frac{\partial s_{ij}}{\partial x_k} dP(\zeta). \tag{20}$$

In practice product demand shares are set by the definition of the outside zero utility composite good. For the logit market share definition drives cross price elasticity estimates. If the outside good share is large, the inside good's shares are small in comparison, generating smaller cross elasticities compared to smaller outside good share. This means that no matter how IIA is relaxed by integrating price response over consumers, price response is still being driven by market share definition and not by product similarity or differences. For the probit equation 17 and 18 demonstrate that market cross elasticity is driven by product similarity or difference. This result makes cross elasticities estimated from the probit demand model robust to definition of the outside good, since they are not directly driven by consumer or market share.

3 Bayesian Econometric Model Specification

Bayesian econometric model specification begins with a hierarchical design. Consider the consumer-level likelihood, the first stage prior and the second stage prior.

- Unit Likelihood $p(y_i|\theta_i)$

- First Stage Prior $p(\theta_i|\tau)$

- Second Stage Prior $p(\tau|h)$

Where θ_i parameterizes the likelihood, τ and h are hyper-parameters characterizing the prior densities of θ and τ respectfully. The unit likelihood is the consumer-level choice model given the data and the modeling parameters. The first stage prior is the random effects distribution, in the present case $P(\zeta)$. The second stage prior forms the econometrician's belief about the distribution of parameters that characterize $P(\zeta)$.

The researcher may believe that a market is made up of many segments. The first stage prior parameters may belong to a distribution that reflects a particular segment of the market. The second stage prior is a belief about the probability process that dictates segment distributions. For example the econometrician can define a mixing process to combine segment distributions generate the market distribution of heterogeneity. A distribution over distributions and their mixing process is referred to as a Dirichlet process. When specifying the second stage prior, the analyst is not forcing the distribution of heterogeneity to have a specific number of segments or modes, the number of components is data driven. The joint posterior for our hierarchical model is given by:

$$p(\theta_1,\ldots,\theta_m,\tau|y_1,\ldots,y_m,h) \propto [\prod_i p(y_i|\theta_i)p(\theta_i|\tau)] \times p(\tau|h)] \tag{21}$$

This specification of the likelihood, $p(y_i|\theta_i)$, demonstrates that the unit-level parameters are only conditionally independent, because τ is given conditional on h.

3.1 Dirichlet Process Prior for Mixing Normals

"Normal densities can be mixed to produce a density of any shape in the same sense that you can build any shape mound by piling up shovels of dirt."(Rossi, Allenby, & McColloch, 2005, p.79) In this spirit N multivariate normal densities may be mixed to form the basis of the distribution of heterogeneity. The

advantage of using a Dirichlet process prior is to identify clusters of data. Consider the following model:

$$
\begin{aligned}
\beta_i &= \Delta\prime z_i + u_i \\
u_i &\sim N(\mu_{ind}, \Sigma_{ind}) \\
ind_i &\sim Multinomial_N(pvec).
\end{aligned}
\tag{22}
$$

ind is a variable that indicates the component distribution that parameter vector i belongs to. ind takes a value from $1, \ldots, K$. $pvec$ is a vector of probabilities that assign weight to each of the normal densities. z_i is a vector of consumer i's demographic characteristics. Δ are coefficients explaining how consumer characteristics relate to marginal utilities, β_i. Moments of β's distribution are computed by,

$$
\begin{aligned}
\mathbb{E}[\beta] &= \sum_k pvec_k \mu_k \\
\mathbb{V}[\beta] &= \sum_k pvec_k \Sigma_k + \sum_k pvec_k (\mu_k - \bar{\mu})(\mu_k - \bar{\mu})).
\end{aligned}
$$

The conjugate prior for the "mixture-of-normals" density is given by:

$$
\begin{aligned}
vec(\Delta) &\sim N(\bar{\delta}, A_\delta^{-1}) \\
pvec &\sim Dirichlet(\alpha) \\
\mu_k &\sim N(\bar{\mu}, \Sigma_k \otimes a_\mu^{-1}) \\
\Sigma_k &\sim IW(\nu_0, V)
\end{aligned}
\tag{23}
$$

N signifies a normal distribution. IW signifies an inverted Wishart distribution.[2] ν_0 is a scale parameter and V is a location and scale parameter. $Dirichlet$ signifies a directed process on the density weighting vector $pvec$ and has a tuning parameter α. This model is applied for the demand systems I estimate.

3.2 MCMC Approach for Bayesian Inference

The Bayesian model proposed in equation 21 does not have a well defined conditional posterior density. As a consequence I implement the random walk Metropolis-Hastings algorithm (Hastings, 1970; Tierney, 1994; Chib & Greenberg, 1995). The Metropolis-Hastings algorithm can, in theory, be applied to any posterior density.[3] The random walk Metropolis-Hastings algorithm performs well on high dimensional problems.

---

[2] For those not familiar with the Wishart density, it is a multivariate generalization of the gamma or chi-squared densities.
[3] A simple way to view this algorithm is much in the spirit of accept/reject sampling. The idea is to accept draws with a probability proportional to the likelihood that they came from the target posterior model.

The cost of using the random walk Metropolis is that is suffers from a higher degree of autocorrelation. This requires the analyst to take close care when selecting the random walk step size and simulating an appropriate amount of draws to adequately summarize the posterior density. A step size that is too large causes the Metropolis to reject draws and consequently does not move. If the step size is too small the researcher will be plagued by a high degree of autocorrelation. The result is that the sampler does not navigate the posterior build-up of mass in the salient regions of the density.

One of the key properties of a random walk is that there is positive probability that the sampler returns to a region of the density infinitely often were the sampler to go on forever. The sampler builds mass by visiting the most probable regions more often. The random walk Metropolis that I implement uses a Gaussian draw function. The draw function is used to make candidate draws. A candidate draw is the old draw plus the shock drawn from the draw function, the multivariate normal density in my case. Rossi et al. (2005) find that the use of covariance values from a first-step maximum likelihood estimate of the the matrix works well and may be adopted as the rule of thumb. The covariance for the draw function may also be tuned with a scaling parameter to ensure that the step size from old to candidate draws is neither too big (causes too many rejections) or too small (does not navigate the entire posterior density).

Metropolis-Hastings algorithms can be used within and along-with Gibbs samplers. Using both is necessary for models I sample from. Rossi et al. (2005) and others verify that hybrid samplers construct chains that have an invariant joint posterior distribution. In this research a Gibbs sampler is used to draw from the distribution of consumer heterogeneity, then at the consumer level I sample from choice models that are estimated for each consumer, using the random walk Metropolis.

The estimation approach exploits the hierarchical structure of the model and makes use of a hybrid sampler. Equations 13 and 14 illustrate that the consumer-level likelihoods for the probit model have parameters characterizing the first moment tied up in the structure of the covariance. Along with the consumer-level scaling parameter, all the parameters in the consumer-level likelihood must be draw at once in a Metropolis-Hastings Step. The component distributions of the density of consumer heterogeneity use a Gibbs sampler because the mixture of normals model that applies a Dirchlet process prior is a conjugate model.

I compute models using the R software package for many reasons. First R is an open source software, so it is available free of charge. For the complex recursive blocks of code, such as the probit likelihood evaluation block, computational bottlenecks are worked out by integrating lower level computer languages. R is well-suited for integrating lower-level computer language. The recursive nature of the simulation means the code is loop intensive. Loop computation is extremely slow in high-level languages like R, so loop-

intensive blocks of the code are written in C, and executed directly at the processor level. The C language communicates directly with the processor, making for efficient computation of loop-intensive blocks, possibly thousands of times faster.[4]

Because the probit model requires one to evaluate integrals over multivariate normal densities the evaluation of the likelihood requires simulation of the integral in equation 3.14. I use the GHK method to approximate the integrals, a method attributable to Keane (1994), Hajivassiliou, McFadden, and Ruud (1996). The GHK methods requires one to truncate normal draws to intervals implied by the multivariate normal density we wish to integrate over. This is equivalent to integrating over rectangular regions of the density. The number of draws required to evaluate the likelihood is 50 to 100. The GHK estimates of the normal probabilities have $\sqrt{n}$ convergence. Consequently increasing the number of draws does not improve estimate accuracy in a significant way. The chance that a probability estimate would cause the wrong choice to be predicted is not likely. The product mistakenly predicted would need to be almost equivalent to the true predicted choice in the eyes of the consumer.[5]

3.3 Evaluation of Model Performance

Bayesian Decision Theory (BDT) is employed to determine which choice model from a set of many is most appropriate for the data at hand. For the current model selection task the action is to determine the "best fitting" choice model or in other words the most probabilistic model. The empirical measure that I use to determine the model with minimal loss is called the log marginal density. One chooses the model for which the log marginal density is largest. Since the metric is a probability, the log transformation results in a negative number, hence chooses the model whose log marginal density is closest to zero.

4   A Sampling Experiment

A Sampling experiment demonstrates the practical behavior of the models and the MCMC samplers that simulate from them for known data generation processes (DGPs). The underlying process that generates data arising from any market is precisely what we intend to capture when we take a demand model to market data for analysis. If one has confidence in the ability on their estimation or simulation routine to recover the DGP then one only needs to evaluate whether their model is reasonable given the data. If MCMC simulators

---

[4]Computer estimation codes are available from the author upon request.

[5]The computer used to execute these routines has a quad core 3.16 GHz processor and 16 GB of RAM. The routine to draw from the posterior of the heterogeneous agent logit requires 90 minutes of computing time to complete fifty thousand draws for a data set consisting of 300 consumers who each have 50 choice sets of three different products. The routine that draws from the posterior of the heterogeneous agent structured probit model takes approximately 3600 minutes, roughly two and a half days.

successfully recover the true densities of the model parameters then they are appropriate for market analysis. This section is divided into two sub-parts. One explains the data experiment I conduct, the other interprets the results.

4.1 Explanation of Experiment

The experiment data is generated from two different processes. The first DGP is a logit model with a distribution of heterogeneity composed by mixing three tri-variate normal components. The second DGP simulates data from a structured covariance probit using distance metric three and a heterogeneity distribution that mixes three tri-variate normal components. The fictional market that generates this data consists of three products. Each products' characteristics are captured by a fixed component that includes unobserved product characteristics, and a variable component that captures an observed product characteristic. Marginal distributions for three specified marketing mix sensitivity parameters reveal consumer marginal preferences for the characteristics. Consumers choose the product that maximizes their utility. This can be expressed as:

$$y_i = max(\beta_{1,i}F_{1,j} + \beta_{2,i}F_{2,j} + \beta_{3,i}C_{1,j}) \text{ for } y \text{ and } j = 1, 2, 3. \tag{24}$$

y is the revealed choice of the consumer. The first heterogeneity parameter $\beta_{1,i}$ has a single mode and is centered about zero. The second heterogeneity parameter, $\beta_{2,i}$, has a marginal distribution that is bimodal. The modes occur roughly at -4 and -1. The third parameter also has a bimodal distribution, its modes occur at -8 and -2.

Fifty purchase occasions for 300 consumers are simulated in this experiment. An example of a single choice outcome for a single consumer appears as

$$y = 3 \qquad X = \begin{bmatrix} 0 & 0 & 2.5 \\ 1 & 0 & 2.9 \\ 0 & 1 & 3.1 \end{bmatrix} \tag{25}$$

where y indicates the choice with an assigned number, and X is a matrix of the characteristics for the options in the choice set. Additionally, 2 characteristics are generated for each consumer. Both consumer characteristics used in the experiment are continuous and imaginably measure income or age.

I specify nine models for analysis. Three different probability models are paired with three different models for the distribution of heterogeneity. The three probability model renditions I evaluate are: The heterogeneous agent logit, and the structured covariance probit applying both distance metrics. The

three models for the distribution of heterogeneity are a one, three, and seven component mixture-of-normal distributions.

All models apply diffuse priors. The priors on the parameters for the logit and probit models are equivalent, but for the θ parameter. Recall that θ_i models the degree of consumer-specific perception of product similarity in the probit error-covariance matrix. There is a prior on the number of normal components, again I will use one, three, and seven components. There is a prior on the mean for each component, these values are set to zero. The prior on the precision matrix of the normal component is 1e-6.[6] The prior parameters for the inverted Wishart, conjugate distribution for the covariance of the heterogeneity parameter, are ν and V. The ν is set to 7 and V is set to $\nu * I_4$. The priors on the Δs in the β regression are mean, $\bar{\Delta}$, and precision matrix, A_d, these are set to zero and 1e-6$*I_4$ respectively. The standard diffuse setting for the alpha parameter, used to tune the Dirichlet Process over the normal components, is set to 5.

Recall from Chapter 4 that MCMC simulation requires us to set various parameter values that tune the sampler. I set the random walk scaling parameter, s, by using a common rule of thumb (Roberts & Rosenthal, 2001).

$$s = \frac{2.93}{\sqrt{nvar + 1}}$$

A fractional likelihood weighting parameter, w, is set to 0.1. This parameter is used to estimate the starting values for the βs. The estimates for the starting values come from maximum likelihood estimates for the logit model. The number of simulated draws for probit likelihood evaluation, r, is set to 50. Increasing the number of simulation draws for the probit likelihood does not increase precision in a significant way since the GHK simulation method has $\sqrt{n}$ convergence. A thinning parameter, keep, is set to 5. This thinning parameter tells the bookkeeping block of the simulation routine to keep only every fifth draw. This practice reduces the autocorrelation inherent in the Random-Walk-Metropolis Sampler. Finally, the number of MCMC draws, R, is set to 50,000 which requires approximately two and a half days of computing time for the probit models and one and a half hours of computing time for the logit model. Computing time can be further reduced on a machine capable of parallel processesing by dividing up the number of draws evenly over the processors. The only problem with this technique is that you also inherent parallel burn-in periods. The burn-in period is amount of draws it takes to free the simulated posterior of influence caused by its initial conditions.

---

[6]The precision matrix is simply the inverse of the prior covariance matrix. The precision matrix is is an information matrix in the classical sense.

4.2 Interpretation of Results

This section presents point estimates for each parameter estimated by each probability model and for a one component heterogeneity distribution. It also presents marginal densities for each parameter estimated by each probability model assuming a single component heterogeneity distribution. Then it displays how well the more flexible, Dirichlet process prior models recover the true marginal densities. Finally it evaluates the fit of each model for both DGPs.

Point Estimates for Single Component Heterogeneity Models

The left side of Table 1 records the posterior marginal mean and joint posterior covariance for the single-component model parameters. All moment estimates are statistically significant in a classical sense. The numerical standard error (n.s.e.) is a normalizing statistic to use if one wishes to conduct a classical hypothesis test. Standard deviation (s.d.) is simply the standard deviation of the posterior parameters over the consumer population. Moment estimates are most appropriate for discussion of single mode marginal posterior densities. Because there are 300 consumers posterior marginal densities, each one is not presented.

Table 1: One Component Posterior Moment Estimates

	Logit DGP				Probit DGP		
	True values	Logit	D2 Probit	D3 Probit	True values	Logit	D2 Probit	D3 Probit
β_1	-0.002	-0.033	0.020	0.022	0.008	-0.200	-0.082	-0.070
s.d.		0.066	0.052	0.052		0.110	0.064	0.065
n.s.e		0.001	0.001	0.001		0.001	0.001	0.001
β_2	-2.448	-2.447	-1.450	-1.454	-2.225	-3.640	-1.845	-1.800
s.d.		0.115	0.063	0.063		0.160	0.070	0.070
n.s.e		0.002	0.001	0.001		0.003	0.002	0.002
β_3	-5.020	-4.906	-3.160	-3.179	-4.502	-6.910	-3.519	-3.550
s.d.		0.190	0.079	0.078		0.280	0.083	0.081
n.s.e		0.003	0.002	0.001		0.005	0.003	0.003
θ	-	-	-3.620	-8.317	0.349	-	0.269	-0.130
s.d.	-	-	0.958	0.892		-	0.117	0.111
n.s.e	-	-	0.114	0.106		-	0.007	0.006
Covariance								
1,1		1.057	0.689	0.697		3.050	1.099	1.125
1,2		-0.078	-0.023	-0.023		-0.420	-0.125	-0.109
2,2		3.062	0.911	0.915		5.720	1.094	1.072
1,3		-0.050	0.002	0.001		0.060	0.034	0.052
2,3		4.012	0.419	0.421		6.510	0.146	0.144
3,3		8.297	1.365	1.378		16.800	1.205	1.184
1,4		-	0.001	0.000		-	0.025	0.048
2,4		-	-0.010	0.001		-	-0.067	-0.041
3,4		-	-0.017	0.000		-	-0.209	-0.222
4,4		-	1.019	1.019		-	1.176	1.186
Model Fit								
LMD		-8088	-8077	-8094		-5993	-5871	-5936
Hit Prob		75.18	77.18	77.18		79.33	83.55	84.00

Source: Authors calculations from sampling experiment

Table 1 also documents estimates of the θ parameter, which is best interpreted after the exponential transformation, vis á vis equation 13. This parameter models the degree of consumer-level model error independence or similarity effect, hence it does not appear in the logit model. Since the data has been generated by a logit model, we expect the exponential transformed value of θ to be small relative to a probit

generated counterpart.

The appropriate test to inform whether the parameter is picking up significant differences in the similarity effect, is to compare the moment estimates of θ from the logit DGP, -3.62 and -8.317, to the estimates of theta from the probit DGP, 0.263 and -0.13. The estimates are clearly larger for the probit DGP by orders of magnitude, which is consistent with greater error dependance across households. This test assumes that the estimates of the β parameter moments are constant across data experiments and experiment estimates. This assumption is reasonable in the current setting since both data sets are generated with the same true βs.

Market-level switching behaviors are accounted for in the structured covariance probitl, as opposed to the logit that suffers from (IIA). However, In the flexible logit model these effects are only accounted for by the distribution of heterogeneity. Therefore, if the structured covariance probit is used on data from a logit generation process, the extent of spread in the true parameter density is potentially dampened since choice probabilities are improperly modeled as a function of error dependence, i.e. the probit is more sensitive then the logit. The moment estimates documented in columns D2 probit and D3 probit of Table 1 reveal that point estimates of the probit parameters are in fact shrunk toward zero, the mean of the prior. Greater prior influence on the posterior distribution of the parameters supports the notion that the information contained in the data is being spread over a larger set of parameters for the probit models. The information is not being lost, it is being used for identification of θ. The more that information from the data is spread for parametric identification, the more influence that information from the prior has on the posterior density.

Posterior Marginal Densities for Single Component Heterogeneity Models

Figure 1 displays the marginal posterior for β_1, under the logit data generation process, over the 300 simulated consumers, for the true distribution and the three model estimates. The distribution of heterogeneity for the models in Figure 1 are specified to have one multivariate normal component, the typical specification for the heterogeneity distribution in the literature. Visually, there is no substantial difference in the estimate for the marginal posterior density of β_1. Table 1 records point estimates for the mean, standard deviation, and covariance of the model parameters for the single component models under both data generation processes.

Next we move on to the multi-modal β_2 parameter. Figure 2 displays marginal posterior densities for the β_2 parameter simulated from each choice model. Each model assumes single component normal heterogeneity. The logit model estimates do a better job at picking up the location of the density, moreover the shoulders of the logit's marginal density estimate contain both modes of the true density. The probit model estimates are similar to each other. They locate the mean of the true density, however they are
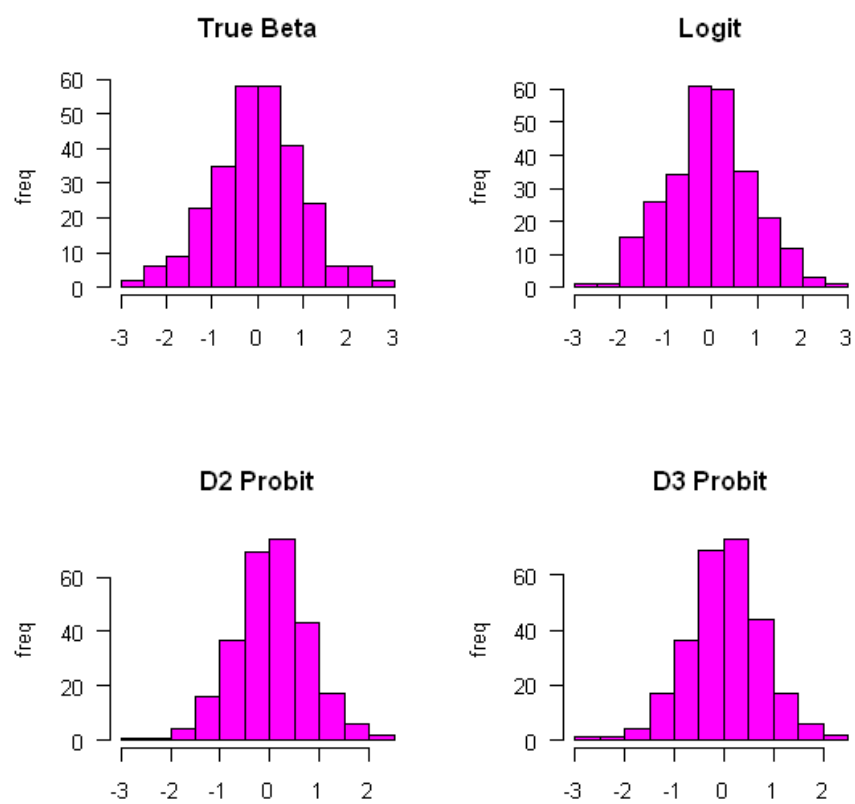
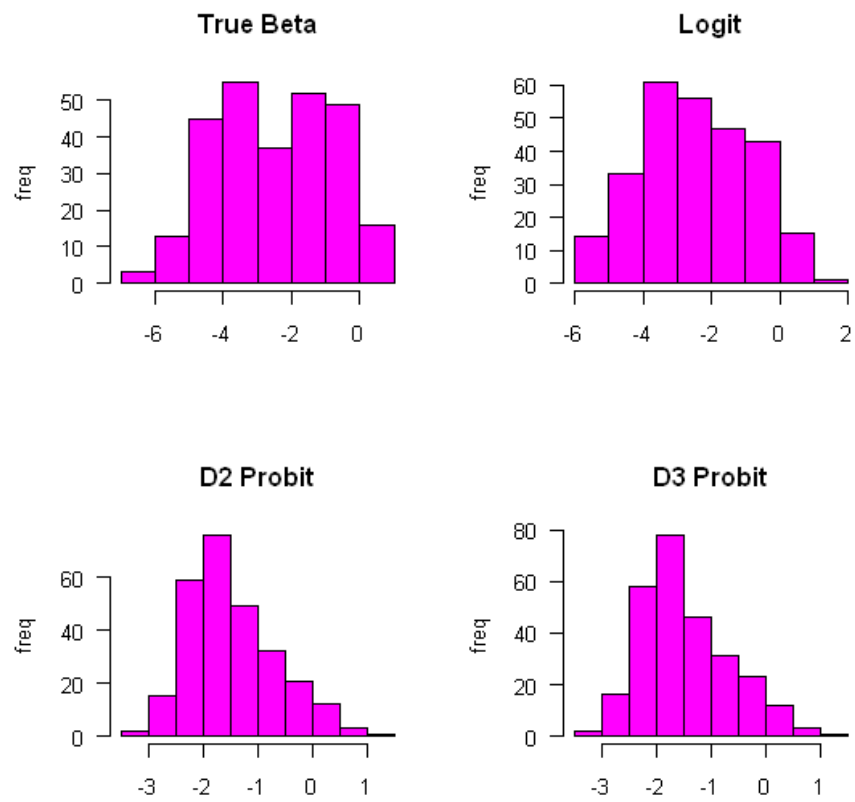18

Figure 1: β_1 Single Component, Logit DGP

Figure 2: β_2 Single Component, Logit DGP

tighter about that mean, suggesting as hypothesized, greater sensitivity of the probit to changes in product characteristics. Also, more Bayesian shrinkage could be occurring in the probit than the logit estimates. Bayesian shrinkage is the tightening of the posterior density. This tightening is attributable to the addition of information in the prior to that already contained in the data and the error model (likelihood). The location of the likelihood moves toward the prior mass and the scale of the likelihood shrinks.

The structured covariance probit possesses an innate ability to predict switching behaviors in the data not attributable to observed choice characteristics. A glance at Figure 3 reveals that the logit model fits the distribution of heterogeneity best. This is possibly due to variation in the consumer distribution capturing similarity effects, whereas it is picked up in the probit model by θ.

Figure 3 displays the marginal densities of β_3 for each model and the true distribution. Once again, we see smaller scale in the marginal densities simulated from the probit models. Shrinkage is evident in the logit results, however not nearly to the extent of the probit results. The moment estimates appearing in Table 5.3 verify these facts. The mean β_3s are clearly closer to zero for the probit models than the logit model, and have smaller deviation. Again we see that the logit model fits the distribution of heterogeneity best.

The exponential transformed marginal posterior densities of θ appear in Figure 4. It is important to keep in mind that θ serves to provided flexibility to the probit covariance matrices with respect to an individual consumer's choice error correlation. Figure 4 indicates that the values of theta for both models are close to zero. We do not expect these distributions to be the same for the two models because they scale the distance metric which is the key differentiating factor between the two models.

Figure 5 plots the marginal densities for the θs simulated under the probit DGP. The true distribution of θs appear at the top. Remember that the probit DGP assumes the second covariance structure. There is a small difference between the θ distribution for the D2 and D3 models. Both do a nice job at picking up the true distribution. If we compare these densities with the densities displayed in Figure 4 we see the values of θ are typically much larger. This is consistent with the point estimates for θ in Table 1. The most important finding in the results about θ is that variation in consumer choices over time provides enough information to identify the distribution of θs. This fact gives us confidence that this model will perform on data that comes from consumer purchase histories.

Analysis of Model Performance on Recovering the Heterogeneity Distribution

Figure 6 displays results from the logit and the D3 probit with 3 and 7 mixing components. The first thing to notice is whether we specify the distribution of heterogeneity to have 3 or 7 components, the posterior appears to display the same distributional features. The reason that the posterior is not forced to pack in all
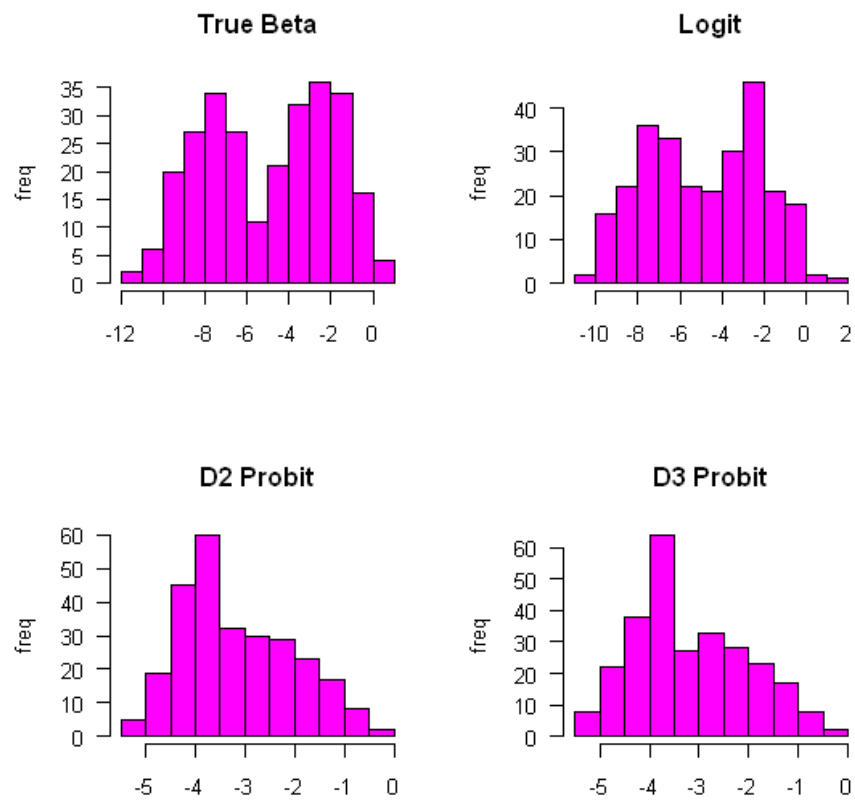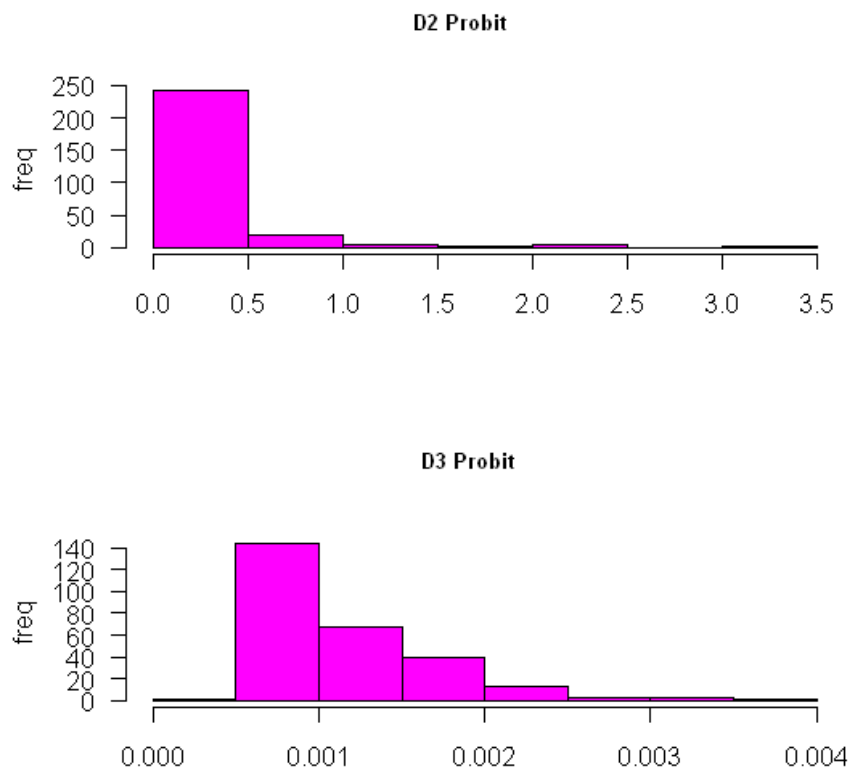
Figure 3: β_3 Single Component, Logit DGP
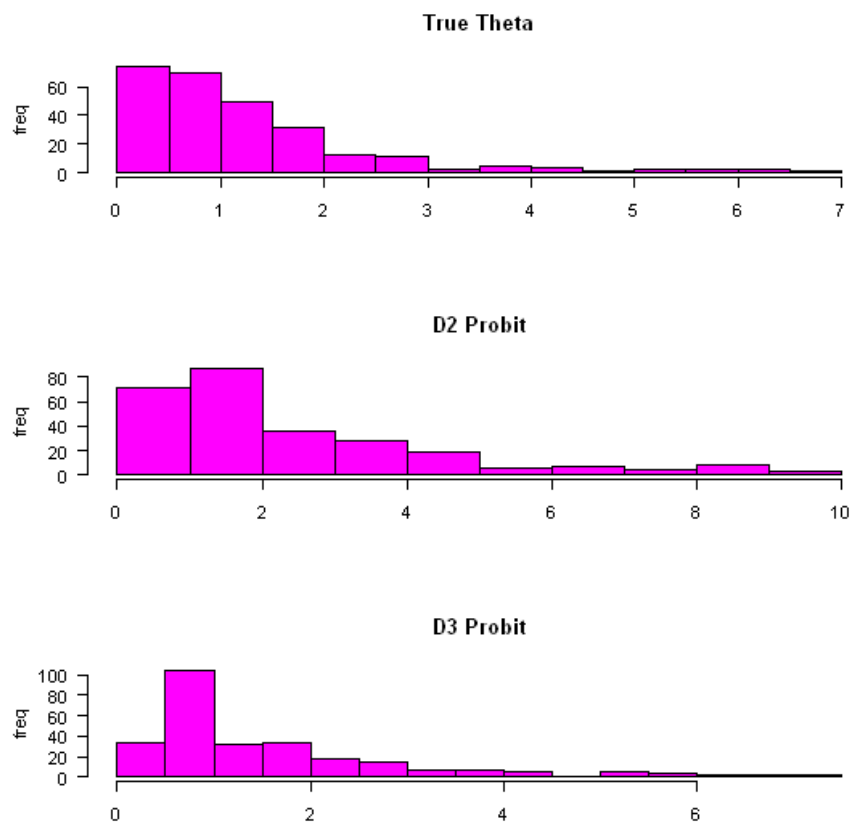
Figure 4: θ Single Component, Logit DGP

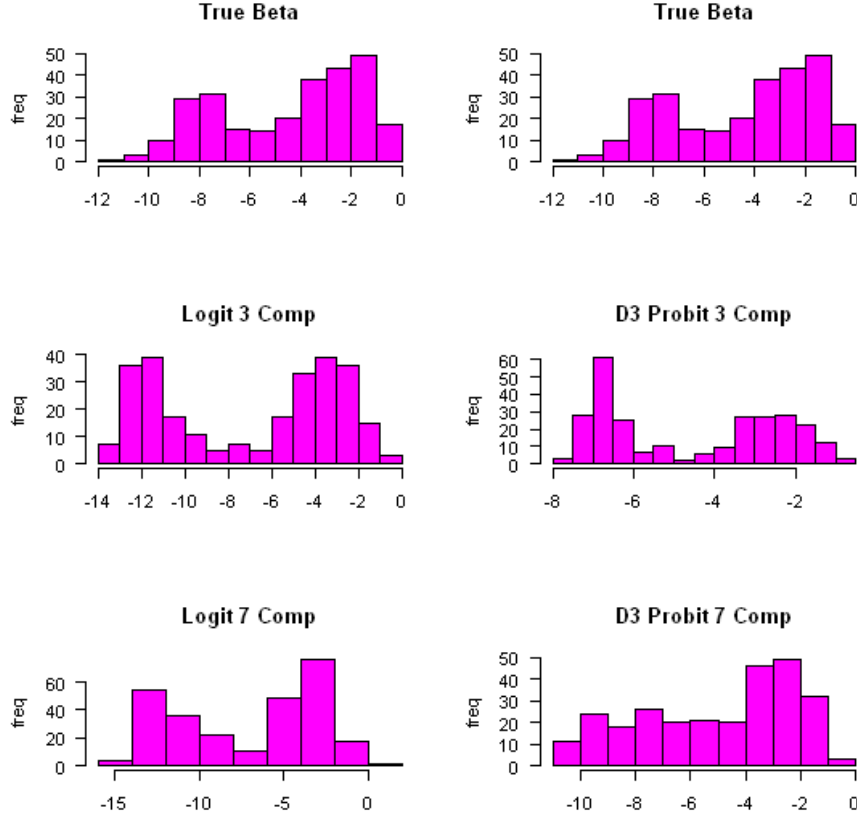Figure 5: θ One Component Probit DGP

Figure 6: Comparison of Component Specifications

7 components is because the simulation process detects only 3 components, and uses just 3 to capture the true joint density. The Dirichlet process effectively places negligible probability weight on 4 of the component densities. This feature of the Dirichlet process prior makes its use very attractive. It lets the data speak for posterior modality rather than dogmatic models that constrain the joint density to a single mode.

Figure 6 also shows that the logit does a slightly better job than the probit models at picking up the distribution of heterogeneity. The logit captures the modality of the True parameter density as well as its spread, its scale is also larger as we would expect. Here we might conclude that the probit seems to be effected by Bayesian shrinkage, although the modes are captured. The middle right diagram of the β_3 distribution has a spike in the left mode. This spike may indicate that the sampler got stuck on a value in that neighborhood, repeating draws because the step size of the random walk was too large. One needs to be careful when taking the model to an application. Various random walk step sizes could be tried, however increasing the number of draws and increasing the value of the thinning parameter does a great deal to improve the performance of a simulation.

25

Model Fit

Table 2 records the Newton-Rafferty estimates of the log-marginal-densities for the models simulated. In choosing the model with the largest (closest to zero) log marginal density estimate it appears that the two probit models outperform the logit almost uniformly. The table also documents hit probabilities for the same models. Results in this table testify that the probit models are superior at predicting the choice outcome even when data is generated from a logit model of choice. The strength of the probit model rests on its ability to model variation in choice by simply adding a single parameter. This is why it outperforms the logit on data generated from a logit DGP. The probit model's superior predictive power is a result of the additional parameter and flexible specification of the error covariance. In the marginal densities we have observed above, the logit does a very nice job at recovering the shape, location, and spread of the parameter distributions. That being said, even the flexible probit did not capture all the salient features of the distribution of heterogeneity for β_3. This result suggests that the probit model fits the data better, while the logit model fits the distribution of heterogeneity better; possibly because information in the data is being used to identify θ.

Table 2: Log Marginal Density and Hit Probabilities

		Logit DGP Normal Componenets			Probit DGP Normal Componenets		
		one	three	seven	one	three	seven
Logit	LMD	-8088	-8076	-8070	-5993	-5996	-5990
	Hit Prob	75.18	75.18	75.18	79.33	79.44	79.42
D2 Probit	LMD	-8077	-7857	-7855	-5871	-5877	-5825
	Hit Prob	77.18	77.14	77.29	83.55	83.47	83.32
D3 Probit	LMD	-8094	-7926	-7987	-5936	-5675	-5927
	Hit Prob	77.18	77.18	77.24	83.72	83.72	83.41

Source: Author's calculations from Sampling Experiment

Table 2 tells the effect a flexible specification of the heterogeneity distribution has on model performance. Across all models and for each DGP, model flexibility improves log marginal density. As for hit probabilities, adding flexibility also improves predictive power. The probit is more flexible than the logit and predicts better without enriching the distribution of heterogeneity. Two important findings about within sample performance of the models can be deduced from these results. One, we see that flexible specification of the logit improves predictive power and two, the probit marginally benefits from the more flexible specification. Adding components results in smaller values for the log marginal density, as Table 2 documents, due to over-parametrization.

Three standards dictate model suitability. One, is the mean and variance from the parameter estimates are attainable, two, is the distribution of consumer heterogeneity recoverable, and three, is the parameter, θ, that models the degree of similarity effects identified by variation in consumer purchase behavior and

consumer characteristics. These standards determine the suitability of the models for demand analysis. Results testify that only models which specify a flexible Dirichlet process prior are suitable for recovering the distribution of heterogeneity. I find that only the two probit models can recover consumer level similarity effects and varying degrees of similarity across the population of consumers. These facts suggest that the two probit models are the most suitable of the three for demand analysis.

5   Empirical Application

To illustrate the properties of each model tested in the sampling experiment I apply them to a real data set on household purchases of lemon-lime carbonated soft drinks. Price elasticities of demand are computed from model estimates to conduct a counterfactual experiment that demonstrates how each model behaves when the market changes. This analysis informs one about how market definition or changes to products in the market impact equilibrium demand sensitivity.

The data is from the New York City Designated Marketing Area (DMA). It records lemon-lime carbonated soft drink purchase histories for a representative sample of households in the metropolitan area. The time frame of the data covers three years beginning in June of 2006 and ending in May of 2009. I select a sub-sample of households purchasing un-cola in the New York City DMA. Un-cola is widely defined as a lemon-lime soda variety. I choose four un-cola products to analyze. These products were by far the post popular during the time period under consideration.[7]

Table 3 records the sample moments for each model assuming a normal-distribution of heterogeneity and the optimal Dirichlet-process distribution of heterogeneity. Once again standard deviations of the sample distributions appear below parameter means and numerical stand errors below them. At the bottom Newton-Rafferty log-marginal-density estimates appear. These guide our model choice.

One will notice that logit parameter estimates are larger in absolute value that probit ones. Recall that this is because the probit model is more sensitive to changes in product characteristics. The reason is that the structured covariance probit model reacts in a chain fashion though the system of choice correlations each depending on consumer sensitivities to characteristics. On the other hand the logit operates with the independent assumption, this means that only a direct effect on the probability is internalized by a change in one of the characteristics, therefore the magnitude of the marginal utility parameters must be larger to capture the total effect.

Since point estimates are less informative about model parameters than sample densities I consider

---

[7]This data has been made available by the Food Marketing Policy Center at the University of Connecticut.

Note: Green bars are 95% confidence intervals, yellow bars are 95% numerical error intervals, and the red bar is the sample mean.
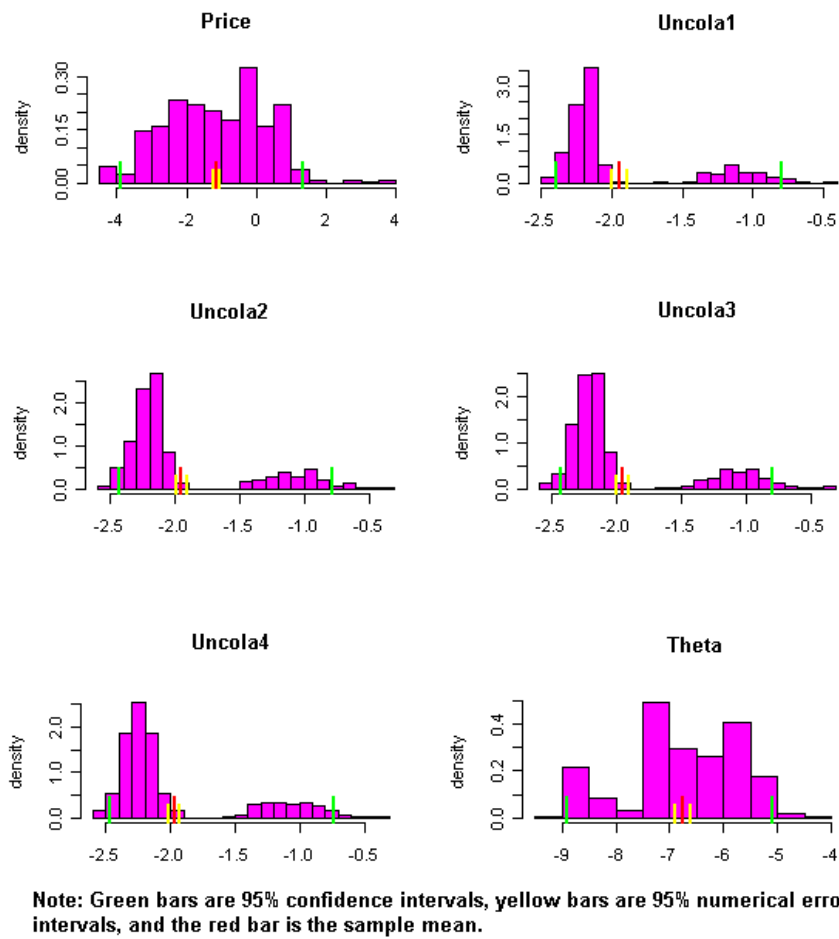
Figure 7: Posterior Marginal Densities for D2 Probit Draws

Price

Uncola1

Uncola2

Uncola3

Uncola4

Theta

Note: Green bars are 95% confidence intervals, yellow bars are 95% numerical error intervals, and the red bar is the sample mean.
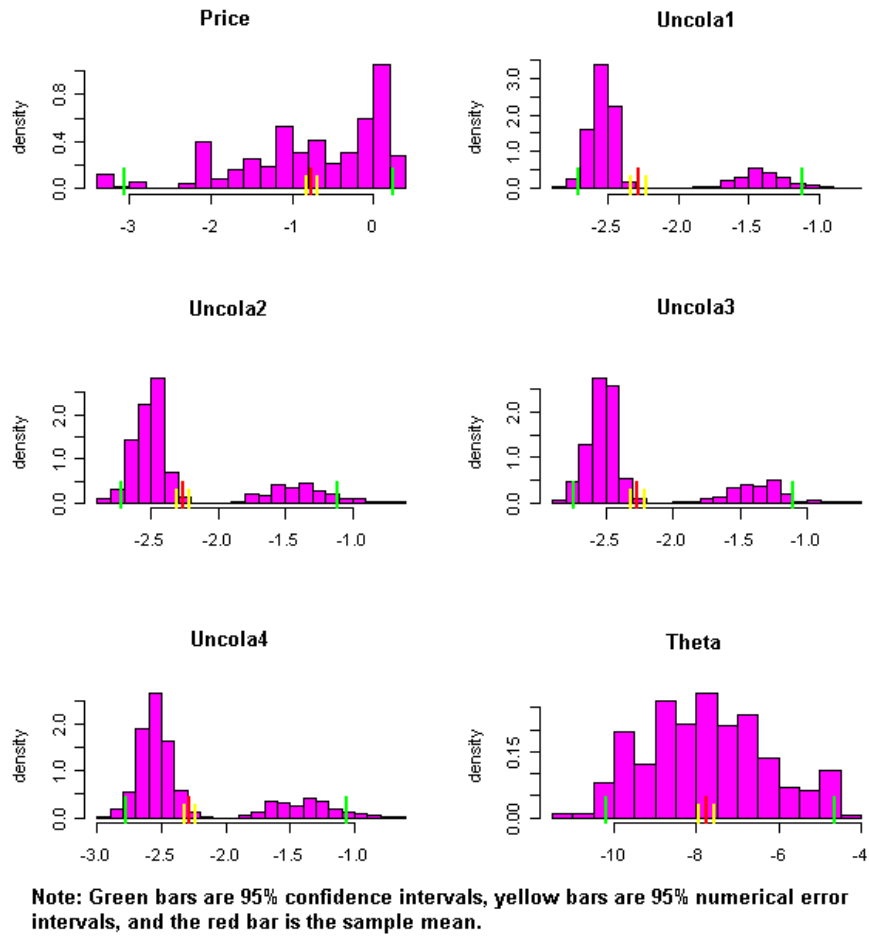
Figure 8: Posterior Marginal Densities for D3 Probit Draws
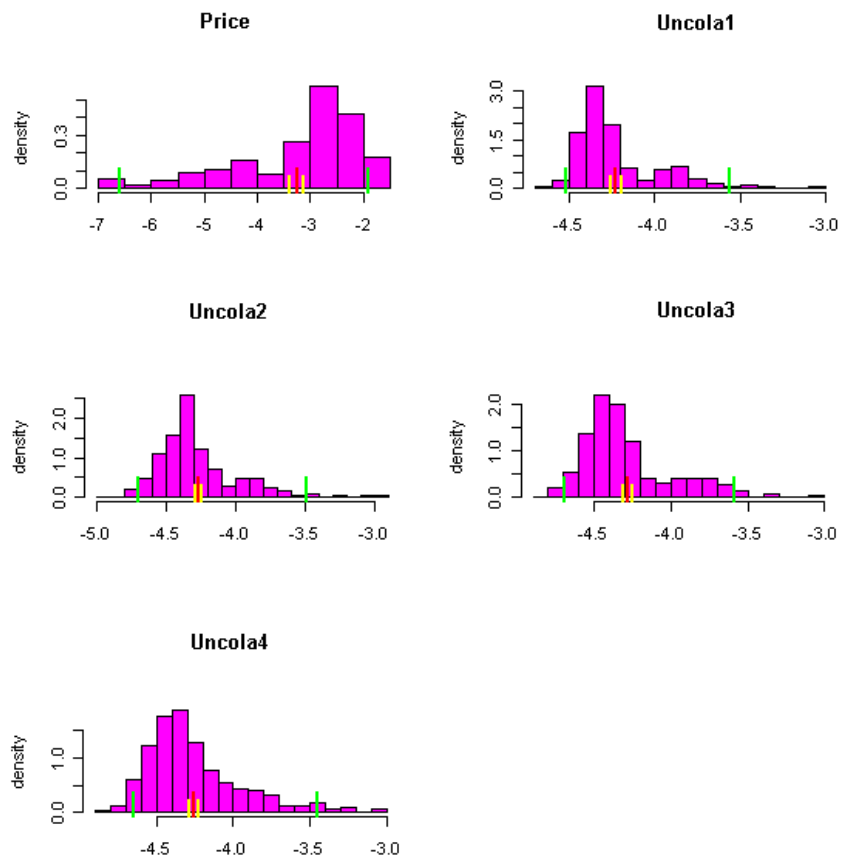
29

Price

Uncola1

Uncola2

Uncola3

Uncola4

Note: Green bars are 95% confidence intervals, yellow bars are 95% numerical error intervals, and the red bar is the sample mean.

Figure 9: Posterior Marginal Densities for Logit Draws

Table 3: Moment Estimates of Posterior Parameter Densities

Parameter	DP comp ▷	Logit		Probit D2		Probit D3	
Models ▷		**One**	**Optimal**	**One**	**Optimal**	**One**	**Optimal**
Price	mean	-3.200	-3.600	-2.053	-1.127	-0.904	-0.760
	s.d.	1.220	3.146	1.373	1.412	0.516	0.880
	n.s.e.	0.064	0.158	0.079	0.045	0.028	0.035
Uncola1	mean	-4.200	-4.837	-2.233	-2.255	-2.254	-2.278
	s.d.	0.260	3.028	0.493	0.492	0.493	0.500
	n.s.e.	0.014	0.136	0.027	0.027	0.027	0.027
Uncola2	mean	-4.300	-4.921	-2.240	-2.264	-2.259	-2.260
	s.d.	0.300	3.479	0.502	0.508	0.505	0.500
	n.s.e.	0.013	0.159	0.023	0.023	0.023	0.022
Uncola3	mean	-4.300	-4.482	-2.243	-2.266	-2.258	-2.620
	s.d.	0.300	1.013	0.515	0.523	0.517	0.510
	n.s.e.	0.013	0.043	0.025	0.025	0.025	0.025
Uncola4	mean	-4.300	-4.698	-2.251	-2.285	-2.262	-2.283
	s.d.	0.320	2.115	0.517	0.530	0.518	0.520
	n.s.e.	0.015	0.096	0.021	0.022	0.022	0.022
θ	mean	-	-	-5.879	-6.750	-9.407	-7.749
	s.d.	-	-	0.650	1.086	1.449	1.470
	n.s.e.	-	-	0.039	0.069	0.090	0.094
	LMD	-3002.9	-2979.6	-2603.8	$-2600.3^{\star}$	-2618.3	-2614.0

Source: Author's calculations
Note: ⋆ indicates the most probabilistic model

density histograms of the parameters from the Dirichlet-process prior models in Table 3. Figures 8-10 display histograms for each of the parameter densities from each model. The densities are labeled according to the variable they measure sensitivity to. These histograms reveal the multi-modality of the parameter densities. The multiple modes identify market segments. They may be interpreted as a segments that are regular un-cola drinkers and a segment that are not. These figures document that very few consumers are actually characterized by the mean parameter since the red bar marks the density mean reported from Table 3. Green bars delineate 95% confidence intervals within which most consumers tastes reside.

Figures 11 and 12 display density estimates of own price elasticities for logit and D2 probit demand models. The top four histograms for each figure show the densities when the data set is made up of households who make ten or more purchases during the sample period. The bottom four histograms for each figure show the densities when the data set is made up of households who made twenty or more purchases during the sample period. These densities reveal no significant difference from their counterpart. This suggests that there are no major systematic differences in the price sensitivity for those who are more regular purchasers, testifying to the robustness of my results and therefore subsequent analysis. Now I use my estimates to analyze the competitive properties of the models.

5.1 Analyzing Market Counterfactuals

In this subsection I examine the validity of the models under consideration for differentiated product demand analysis. Most prominently I consider how implementation of a consumer-level logit model, that suffers from
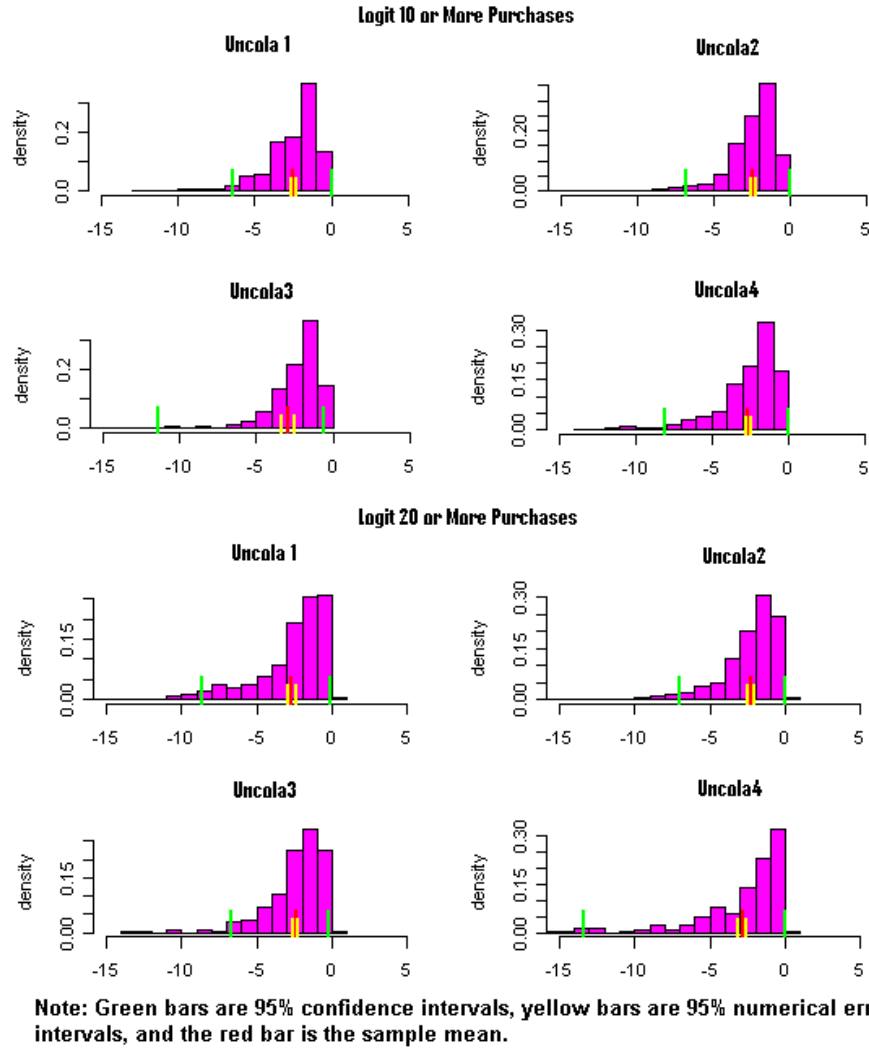
Figure 10: Own Price Elasticity Density Estimates for Logit: For Data Selected of Households with Ten and Twenty or more purchases

Note: Green bars are 95% confidence intervals, yellow bars are 95% numerical error intervals, and the red bar is the sample mean.
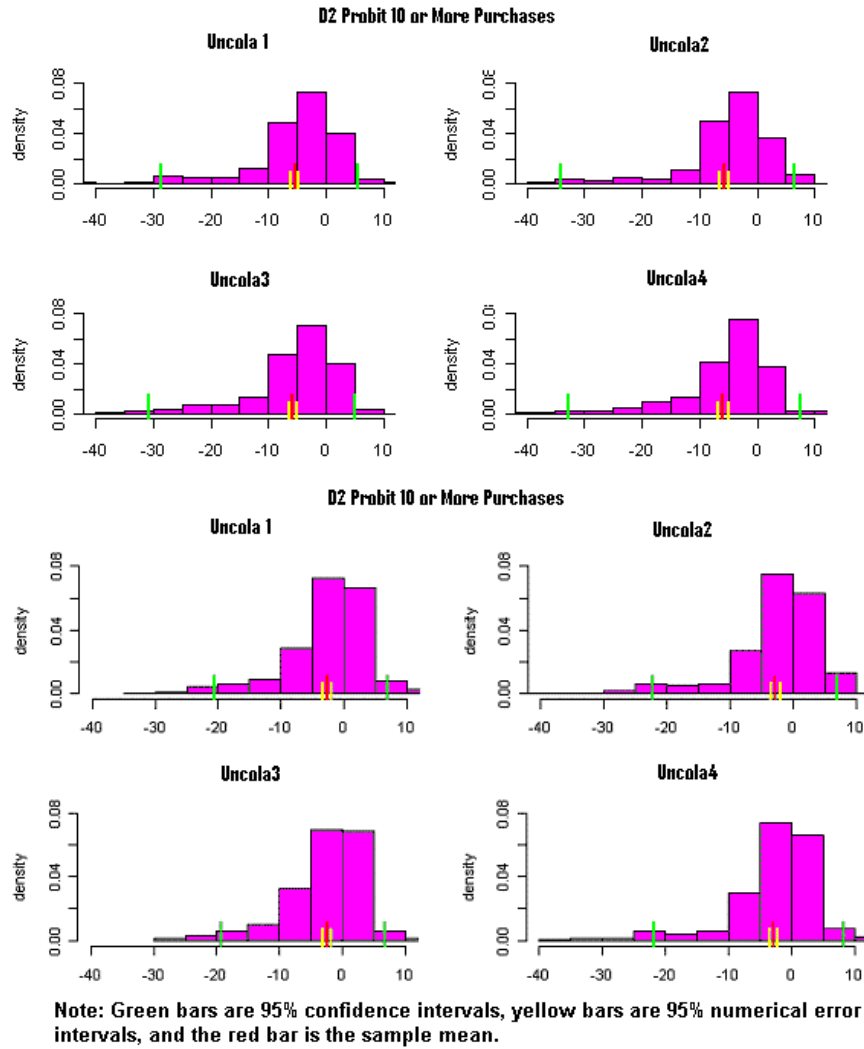
Figure 11: Own Price Elasticity Densities for D2 Probit: For Data Selected of Households with Ten and Twenty or more purchases

the IIA property, influences measures of competition implied by demand estimates. I also provide arguments for adopting the structured-covariance probit instead. First I explain a test of product market definition. Then I present a simple counterfactual that reveals how logit and probit demand models determine price elasticity before and after a multi-lateral perturbation in utility from inside good consumption. Then I analyze the results of the experiment and comment on the use of each model for competitive analysis.

To validate each model I begin with a hypothesis about how price sensitivity should change when outside good share is reduced due to a multi-lateral equi-sized fixed increase in the intercept value of the inside goods. This is equivalent to augmenting the product market by taking away a good that is not in competition with inside goods. For instance assume I take way the option of purchasing a bicycle as an outside good. Bicycles are a high value choice and one has no reason to believe they are in competition with un-colas. I would be taking away a more valuable good, effectively making inside goods more valuable relative to the outside options, which no longer include a bicycle. As such our demand model should yield inside good cross price elasticity estimates that are robust to market definitions that include irrelevant outside goods.

Given that the value of inside goods increases relative to the outside option, robustness implies the following two conditions be met. First, since the goods are now uniformly more valuable they will become less elastic. Second, since the relative value of inside goods has not changed there should be only a small reduction in cross-price sensitivity solely because their respective demand curves become steeper. This establishes two standards for judging model estimates of price sensitivity. One, cross-price elasticities should be marginally smaller, and two, own price elasticities should become significantly larger.

Tables 4-6 report price elasticity matrices from each probit, and the logit demand models before and after a multi-lateral fixed increase in inside good utility across consumers. In Tables 4 and 5 one will observe that the two standards established above are met. Own price elasticities are larger across all products and Cross price elasticities are slightly smaller.

Table 6 reports the experiment's outcome for the logit model. Here one will notice that cross-price elasticities are higher and own-price elasticities are only slightly larger. This attests to an important fact. IIA dictates that switching is proportional to market share and introduction of heterogeneity does not completely overcome it. Because market share increased for inside goods relative to the outside goods, share went up across the board. It should be no surprise that cross-price elasticities went up since they are a multiplicative function of share. This fact is not overcome by introducing a distribution of heterogeneity. Results for Tables 4-6 suggest that structured covariance probit models are more robust to specification of the outside good than logit models are. Logit cross-price effects are very sensitive to a market definition and therefore they

Table 4: Market Definition's Effect on Probit Price Elasticities: A Test of IIA

| | Before Unilateral Utility Increase | | | |
Model: D3 Probit	Uncola1	Uncola2	Uncola3	Uncola4
Uncola1	-3.55	0.3546	0.357	0.406
Uncola2	0.342	-3.591	0.347	0.385
Uncola3	0.331	0.336	-3.639	0.403
Uncola4	0.394	0.39	0.423	-4.076
Outside	0.04	0.042	0.045	0.048

Source: Author's calculations

| | After Unilateral Utility Increase | | | |
Model: D3 Probit	Uncola1	Uncola2	Uncola3	Uncola4
Uncola1	-2.853	0.323	0.326	0.368
Uncola2	0.309	-2.842	0.314	0.346
Uncola3	0.302	0.307	-2.905	0.364
Uncola4	0.363	0.36	0.387	-3.293
Outside	0.048	0.051	0.054	0.057

Source: Author's calculations
Note: Change is for share in a row against the price in a column

should be used with great caution.

Results from Tables 4 and 5 verify that structured covariance probit models are most appropriate for determination of relevant product markets. Therefore any competitive analysis based on estimates of demand should adopt a structured covariance probit modeling approach in lieu of a logit one. The structured covariance probit's ability to model similarity effects, can retain stable estimates of inside good price sensitivity regardless of goods included in the outside analysis. This is because under many market specifications inside good probabilities are from the lower tail of the distribution function determining share, i.e. goods have small market share relative to the outside option. The structured covariance probit's behavior is more stable than that of the logit in the tails because it accounts for covariance in choice.

These findings suggest that structured-covariance probit models do well in determining whether goods belongs to a product market. For example, goods can be added or taken away from the analysis and the model will use its structure to determine whether the presence of the good in question raises equilibrium price. This type of analysis is common for market definition in antitrust proceedings. These tests of relevant product market definition rely on determining whether a small but significant non-transitory increase in price (SSNIP) is profitable for the hypothetical monopolist. If such an increase is not profitable more goods are removed from the analysis until the price increase becomes profitable. This type of analysis relies on dependable elasticity estimates. Results from this subsection suggests that the heterogeneous-consumer logit

Table 5: Market Definition's Effect on D2 Probit Price Elasticities: A Test of IIA

| Model: D2 Probit | Before Unilateral Utility Increase | | | |
	Uncola1	Uncola2	Uncola3	Uncola4
Uncola1	-5.293	0.517	0.527	0.569
Uncola2	0.504	-5.399	0.594	0.556
Uncola3	0.511	0.592	-5.651	0.577
Uncola4	0.564	0.566	0.590	-5.850
Outside	0.061	0.064	0.068	0.069

Source: Author's calculations

| Model: D2 Probit | After Unilateral Utility Increase | | | |
	Uncola1	Uncola2	Uncola3	Uncola4
Uncola1	-3.944	0.457	0.467	0.500
Uncola2	0.455	-4.091	0.534	0.499
Uncola3	0.447	0.514	-4.120	0.500
Uncola4	0.495	0.496	0.516	-4.303
Outside	0.078	0.082	0.087	0.088

Source: Author's calculations

should never be used for such an analysis and that a more reasonable modeling approach would be to use the structured-covariance probit.

Results from Table 3 indicate that the D2 Probit demand model with a flexible Dirichlet-process prior was the most probabilistic model making it the favored one in a Bayesian decision analysis. The D3 Probit also fared well. Results from the Bayesian decision analysis also testify that including the Dirichlet-process prior to flexibly model the distribution of heterogeneity improves all three models. Numerical standard errors, sampler draw- and autocorrelation function- plots indicated that samplers for each of the models performed well.

The probit models offered up the most realistic option as a demand models. They posses the most realistic patterns of substitution unhindered by IIA, as in the logit. This result was bolstered by estimates of a rich distribution of varying consumer perception to the degree of product differentiation (or similarity), driven by the θ parameter. This result is consistent with the hypothesis that consumers vary in the way they differentiate products. A flexible model like the structured covariance probit appreciates the differences whereas the logit and any demand system that views products as symmetrically differentiated does not. Finally, the probit models appear to be more robust to definition of the outside choice, further supporting the case for adopting the probit model for demand analysis. I now move to the next chapter where I summarize the findings of the research and make suggestions for future extensions of it.

Table 6: Market Definition's Effect on Logit Price Elasticities: A Test of IIA

| | Before Unilateral Utility Increase | | | |
Model: Logit	Uncola1	Uncola2	Uncola3	Uncola4
Uncola1	-3.22	0.043	0.043	0.048
Uncola2	0.043	-3.116	0.041	0.044
Uncola3	0.044	0.041	-3.151	0.046
Uncola4	0.048	0.043	0.045	-3.342
Outside	0.044	0.042	0.042	0.046

Source: Author's calculations

| | After Unilateral Utility Increase | | | |
Model: Logit	Uncola1	Uncola2	Uncola3	Uncola4
Uncola1	-3.155	0.106	0.105	0.119
Uncola2	0.107	-3.05	0.1	0.108
Uncola3	0.107	0.101	-3.08	0.113
Uncola4	0.118	0.106	0.11	-3.26
Outside	0.109	0.104	0.104	0.113

Source: Author's calculations

6 Conclusion

This paper introduced a new structured covariance probit demand model. This model fully relaxes the independence of irrelevant alternatives property inherent in logit choice models. It does so by modeling choice covariance with a structural formulation that relies on the utility differences. The Dirichlet-process prior modeled - multi-modal, skewed, and heavy tailed - distributions of consumer tastes. A hybrid sampler recovered the posterior density of the structured covariance probit demand system parameters. The sampling experiment confirmed that the models and algorithms recover the posterior distribution of a known data generation process successfully. An empirical application to the New York City un-cola market demonstrated the structured covariance probit's ability to estimate consumer switching behaviors in comparison to the popular heterogeneous-consumer logit. Findings here certify that the logit model should be used with caution, particularly as caution applies to specification of the outside good. Future extensions of this research might explore alternate structures of the choice correlation model or develop an estimation approach to use the new model for examination of aggregate data. Empirical analysis could apply the structured covariance probit model to questions about market definition, the introduction of new goods, optimal advertising, and merger analysis.

References

Ackerberg, D., Benkard, C., Berry, S., & Pakes, A. (2007). *Econometric Tools for Analyzing Market Outcomes*, pp. 4171–4276. Amsterdam: North-Holland. eds: James J. Heckman and Edward E. Leamer.

Albert, J., & Chib, S. (1993). A Bayesian Analysis of Binary and Ploychotomous Response Data. *Journal of the American Statistical Association*, 669–679.

Allenby, G. (1989). A Unified Approach to Identifying, Estimating and Testing Demand Structures with Aggregate Scanner Data. *Marketing Science*, *8*(3), 265–280.

Allenby, G., & Rossi, P. (1991). There Is No Aggregate Bias: Why Macro Logit Models Work. *Journal of Business and Economic Statistics*, *9*(1), 1–14.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, *63*, 841–889.

Berry, S. T. (1994). Estimating Descrete Choice Models of Product Differentiation. *RAND Journal of Economics*, *25*(2), 242–262.

Bunch, D. (1991). Estimibility in the Multinomial Probit Model. *Transportation Research Part B*, pp. 1–12.

Burda, M., Harding, M., & Hausman, J. (2008). A Bayesian Mixed Logit - Probit Model for Multinomial Choice. *Journal of Econometrics*, *147*(2), 232–246.

Carroll, J., & Green, P. (1995). Guest Editorial: Psycometric Methods in Marketing Research: Part i, Conjoint Analysis. *Journal of Marketing Research*, 385–391.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *American Statistician*, 327–335.

Chintagunta, P., Jain, D., & Vilcassim, N. (1991). Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data. *Journal of Marketing Research*, *28*, 417–428.

Debreu, G. (1960). Review of R. D. Luce Individual Choice Behavior. *American Economic Review*, *50*, 186–188.

Dube, J., Hitsch, G., & Rossi, P. (2009). State Dependence and Alternative Explanations for Consumer Inertia. *NBER Working Paper*.

Graybill, F. (1976). *Theory and Application of the Linear Model*. Duxbury Press: North Scituate MA.

Hajivassiliou, V., McFadden, D., & Ruud, P. (1996). Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives. *Journal of Econometrics*, *72*(1), 85–134.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, *57*(1), 97–109.

Hausman, J., & Wise, D. (1978). A Conditional Probit Model for Qualitative Choice: Discrete Decisions Decognizing Interdependence and Heterogeneous preferences. *Econometrica*, 403–426.

Hofacker, C. (1990). Derivation of Covariance probit Elasticities. *Management Science*, 500–504.

Keane, M. P. (1994). A Computationally Practical Simulation Estimator for Panel Data. *Econometrica*, 95–116.

Lenk, P., Dotson, J., Brazell, J., Otter, T., Maceachern, S., & Allenby, G. M. (2009). A Probit Model with Structured Covariance for Similarity Effects and Source of Volume Calculations. Working paper, Fisher College of Business, Ohio State University.

McColloch, R., & Rossi, P. E. (1994). An Exact Likelihood Analysis of the Multinomial Probit Model with Fully Identified Parameters. *Journal of Econometrics*, pp. 207–240.

McColloch, R. E., Polson, N., & Rossi, P. E. (2000). A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Paramters. *Journal of Econometrics*, pp. 173–193.

McFadden, D. (1981). Econometric Models of Probabilistic Choice. *Structural Analysis of Discrete Data with Econometric Applications*, *198272*.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, *15*(5), 447–470.

Poon, W., & Lee, S. Y. (1987). Maximum Likelihood Estimation of Multivariate Polyserial and Polychoric Correlation Coefficients. *Psycometrika*, *52*(3), 409–430.

Roberts, G., & Rosenthal, J. (2001). Optimal Scaling for Various Metropolis-Hastings Algorithms. *Statistical Science*, 351–367.

Rossi, P. E., Allenby, G. M., & McColloch, R. E. (2005). *Bayesian Statistics in Marketing*. John Wiley and Sons, Ltd.

Tchumtchoua, S. (2008). *Advertising and Dynamic Demand for Differentiated Products*. Ph.D. Dissertation, University of Connecticut.

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *the Annals of Statistics*, 1701–1728.

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

Tversky, A. (1972). Elimination by Aspects. *Psychological Review*, *79*(4), 281–301.

Tversky, A. (2003). Elimination by Aspects: A Theory of Choice. *Preference, Belief, and Similarity: selected writings*, 463.

Yai, T., Iwakura, S., & Morichi, S. (1997). Multinomial Probit with Structured Covariance for Route Choice Behavior. *Transportation Research Part B*, *31*(3), 195–207.

A   Appendix: Probit Elasticity Derivation

Derivation of the probit elasticity follows Hofacker (1990). First I define difference utilities, then I show the form of the probit derivative with respect to a choice characteristic.

A.1   Defining Difference Utilities

The probability that product i is chosen is:

$$
\begin{aligned}
S_i &= Prob[u_i > u_j] \\
&= Prob[u_i - u_j > 0] \\
&= Prob[\nu_j^{(i)} > 0].
\end{aligned}
\tag{26}
$$

In the third line of this equation I define $u_i - u_j = \nu_v^{(i)}$ therefore I will introduce a difference operator for each brand, I call it $\Lambda^{(i)}$. This operator subtracts each other brands from the ith. It is a n by N matrix where N is the number of products and $n = N - 1$. For example, $\Lambda^{(1)}$ is:

$$
\Lambda^{(1)} = \begin{bmatrix}
1 & -1 & 0 & \cdots & 0 \\
1 & 0 & -1 & \cdots & 0 \\
& \cdots & & \cdots & \cdots \\
1 & 0 & 0 & \cdots & -1
\end{bmatrix}
$$

Apply this operator to u to produce $\nu^{(i)} = \Lambda^{(i)} u$. This vector is normally distributed based on the model error assumption with:

$$E[\nu^{(i)}] \equiv \hat{\nu}^{(i)} = \Lambda^{(i)} \hat{u} = \Lambda^{(i)} x' \beta$$

$$V[\nu^{(i)}] \equiv \Psi^{(i)} = \Lambda^{(i)} \Sigma \Lambda^{(i)\prime}$$

Standardize $\nu^{(i)}$. Define $\Delta^{(i)} = diag[\Psi^{(i)}]$, then transform the covariance matrix of difference utilities.

$$R^{(i)} = [\Delta^{(i)}]^{-1/2} \Psi^{(i)} [\Delta^{(i)}]^{-1/2},$$

and expectation of difference utilities

$$a^{(i)} = [\Delta^{(i)}]^{-1/2} \hat{\nu}^{(i)}.$$

A.2 Probit Derivatives w.r.t. Product Characteristics

Differentiating the probit probability with respect to product charecteristics can be broken into two parts according to the chain rule as:

$$\frac{\partial s_i}{\partial x'} = \frac{\partial s_i}{\partial u^{(i)}{}_\prime} \frac{u^{(i)}}{\partial x'}$$

The $\frac{u^{(i)}}{\partial x'}$ is obviously β. The $\frac{\partial s_i}{\partial u^{(i)\prime}}$ is a bit less obvious and takes the form, Poon and Lee (1987) show:

$$\frac{\partial s_i}{\partial u_j^{(i)}} = \phi(u_j^{(i)}) \times \Phi[\cdots, \frac{u_k^{(i)} - r_{jk}^{(i)} u_j^{(i)}}{(1 - [r_{jk}^i]^2)^{1/2}}, \cdots ; R_{\cdot j}^{(i)}], \quad k \neq j.$$

Where:

$$R_{\cdot j}^{(i)} = [\Delta_{\cdot j}^{(i)}]^{-1/2} \Psi_{\cdot j}^{(i)} [\Delta_{\cdot j}^{(i)}]^{-1/2},$$

and the $\cdot j$ operator indicates that row and column j are left out of the matrix. This approach calculates $\Psi_{\cdot j}^{(i)}$ under the assumption that $\nu^{(i)}$ has been held constant, or "partialed out."(Graybill, 1976)

Then we measure own response:

$$\frac{\partial s_i}{\partial x_i} = \beta \sum_{j=1}^{n} \frac{\partial s_i}{\partial u_j^{(i)}} [\delta_{jj}^{(i)}]^{-1/2}, \tag{27}$$

and cross response

$$\frac{\partial s_i}{\partial x_j} = -\beta \frac{\partial s_i}{\partial u_j^{(i)}} [\delta_{jj}^{(i)}]^{-1/2} \tag{28}$$

That completes the derivation.∎

B   Appendix: Description of Data Used in the Empirical Analysis

In this appendix I present descriptive statistics on the variables used in demand analysis for un-cola consumers in the New York City DMA. These variables include demographics, market prices, and market brand choice shares. First I present a table summarizing demographic variables. Then I present a table summarizing prices and market brand shares for the four un-colas under considers in my demand study. I consider households that purchased un-cola ten or more times during the three year period. I analyze the robustness of demand results by estimating the best models with a data set that records transactions for households that purchased un-cola more than twenty times. This test of robustness verifies whether there is selection bias resulting from a systematic difference in the purchase shares for more regular consumers of un-cola.

Table 7: Descriptive Statistics for Demographics used in Analysis

Variable		Mean	Std. Dev.	Min	Max
Income		58k	35k	5k	110k+
Family Size		2.81	1.34	1	9
Race		**White**	**Black/African**	**Asian**	**Other**
	Households	360	84	24	41
	Percent	70.73	16.5	4.72	8.06
	Total Households	509			

Source: A.C. Nielsen Home-scan

Table 5.7 provides descriptive statistics for household demographic characteristics used to model the distribution of parameters measuring marginal dis-utility of income and tastes for products. Their are 509 households under consideration in my analysis. The average household incomes range from 5,000 to 110,000+ dollars and the average is roughly 58,000 dollars. The average household has 2.81 members. The race in most households is white with black/African coming in distant second.

Table 5.8 documents descriptive statistics of prices and purchase shares for the un-colas I analyze. The prices are for 12oz servings, the size of typical soda can. Shares are simply the proportion of times the product was purchased during the three years observed. These are not volume or unit market shares. They should be considered brand choice market shares. Marketers are interested in brand choice because is reflects the proliferation of their product and sensitivity of product adoption to promotion and pricing strategy. As such, marketing strategy is often shaped by the brand choice share metric, I cite as evidence the popular conjoint analysis field of marketing research (Carroll & Green, 1995). Conjoint analysis is a

statistical technique used in marketing research to determine how people value different features that make up a product based on choice outcomes.

Table 8: Descriptive Statistics for Price and Brand Share

Variable	Product	Mean	Std. Dev.	Min	Max
Price (cents per 12oz)					
	Uncola1	0.253	0.359	0.012	2.621
	Uncola2	0.192	0.295	0.012	1.628
	Uncola3	0.160	0.215	0.014	1.098
	Uncola4	0.167	0.298	0.007	2.090
	Market	0.193	0.292	0.007	2.621
Share		Count	Inside Share	Total Share	
	Uncola1	157	25.2	1.73	
	Uncola2	157	25.2	1.73	
	Uncola3	154	24.72	1.69	
	Uncola4	155	24.88	1.71	
	Outside Choice	8,464		93.14	
	total	9087			

Source: A.C. Nielsen Home-scan

In Table 5.8 one will note that Uncola1 is the highest priced product at approximately 25 cents per can and leads in the number of purchase occasions along with Uncola2 at 157 times. Uncola3 is the lowest priced cola and purchased the fewest number of times. The outside choice is defined as any carbonated soft drink purchase made by a household under consideration. This choice is a zero utility choice. All utility parameters are identified relative to the outside carbonated soft drink option. Next I present results from models estimated with this data.

FOOD MARKETING POLICY CENTER
RESEARCH REPORT SERIES

This series includes final reports for contract research conducted by Policy Center Staff. The series also contains research direction and policy analysis papers. Some of these reports have been commissioned by the Center and are authored by especially qualified individuals from other institutions. (A list of previous reports in the series is available on our web site.) Other publications distributed by the Policy Center are the Working Paper Series, Journal Reprint Series for Regional Research Project NE-165: *Private Strategies, Public Policies, and Food System Performance*, and the Food Marketing Issue Paper Series. Food Marketing Policy Center staff contribute to these series. Individuals may receive a list of publications in these series and paper copies of older Research Reports are available for $20.00 each, $5.00 for students. Call or mail your request at the number or address below. Please make all checks payable to the University of Connecticut. Research Reports can be downloaded free of charge from our web site given below.