Management Science

A Bayesian Semiparametric Approach for Endogeneity and Heterogeneity in Choice Models

Yang Li, Asim Ansari

A Bayesian Semiparametric Approach for Endogeneity and Heterogeneity in Choice Models

Yang Li

Cheung Kong Graduate School of Business, Beijing 100738, China, yangli@ckgsb.edu.cn

Asim Ansari

Columbia Business School, Columbia University, New York 10027, maa48@columbia.edu

Marketing variables that are included in consumer discrete choice models are often endogenous. Extant treatments using likelihood-based estimators impose parametric distributional assumptions, such as normality, on the source of endogeneity. These assumptions are restrictive because misspecified distributions have an impact on parameter estimates and associated elasticities. The normality assumption for endogeneity can be inconsistent with some marginal cost specifications given a price-setting process, although they are consistent with other specifications. In this paper, we propose a heterogeneous Bayesian semiparametric approach for modeling choice endogeneity that offers a flexible and robust alternative to parametric methods. Specifically, we construct centered Dirichlet process mixtures (CDPM) to allow uncertainty over the distribution of endogeneity errors. In a similar vein, we also model consumer preference heterogeneity nonparametrically via a CDPM. Results on simulated data show that incorrect distributional assumptions can lead to poor recovery of model parameters and price elasticities, whereas the proposed semiparametric model is able to robustly recover the true parameters in an efficient fashion. In addition, the CDPM offers the benefits of automatically inferring the number of mixture components that are appropriate for a given data set and is able to reconstruct the shape of the underlying distributions for endogeneity and heterogeneity errors. We apply our approach to two scanner panel data sets. Model comparison statistics indicate the superiority of the semiparametric specification and the results show that parameter and elasticity estimates are sensitive to the choice of distributional forms. Moreover, the CDPM specification yields evidence of multimodality, skewness, and outlying observations in these real data sets.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mnsc.2013.1811.

Keywords: discrete choice; endogeneity; semiparametric bayesian; centered dirichlet process mixtures; heterogeneity
History: Received July 18, 2011; accepted May 27, 2013, by J. Miguel Villas-Boas, marketing. Published online in *Articles in Advance* December 23, 2013.

1. Introduction

Over the past decade, a growing number of studies have documented the importance of accounting for endogeneity and heterogeneity in discrete choice models involving aggregate (e.g., Berry et al. 1995, Chintagunta 2001, Park and Gupta 2009) or disaggregate choice data (e.g., Chintagunta et al. 2005). Prices and other marketing variables are often endogenous because these are set by firms taking into account product attributes that are unobserved by the researcher. This results in a correlation between the observed marketing variables that are included in the systematic component of utility functions and the unobserved factors. It is well known that failure to account for the endogeneity of marketing variables leads to inconsistent parameter estimates (Villas-Boas and Winer 1994, 1999). Similarly, a failure to account for individual differences in model parameters can yield misleading inferences about consumer response

sensitivities. Both types of inferential problems can have important consequences for managerial actions.

A number of different approaches have been proposed for handling the endogeneity problem in individual-level discrete choice models. These range from structural approaches that explicitly model the supply side using a game (e.g., Yang et al. 2003, Villas-Boas and Zhao 2005) to limited information approaches that model the price-setting process as a linear equation (e.g., Villas-Boas and Winer 1999). The latter can be considered a "reduced-form representation" of an underlying supply-side model. A variant of the limited information approach is the recently proposed control-function method (Petrin and Train 2010) that uses extra variables to control for the portion of the variation in the unobserved factors that is not independent of prices. Endogeneity is also handled using brand and time-specific fixed effects in the utility function. These fixed effects

represent unobserved attributes of brands that are correlated with prices (Goolsbee and Petrin 2004, Chintagunta et al. 2005). A number of estimation methods have been used in dealing with the endogeneity problem. These include generalized method of moments (GMM), maximum likelihood estimation, fixed-effects, and two-step approaches as well as Bayesian methods (Chintagunta et al. 2005, Yang et al. 2003, Rossi et al. 2005, Kuksov and Villas-Boas 2008).

In this paper, we investigate how inferences about model parameters and price elasticities in *individual-level* discrete choice models are sensitive to the distributional assumptions about endogeneity and heterogeneity errors. We study whether misspecification of these distributional forms matters and propose a heterogeneous Bayesian semiparametric approach for simultaneously modeling endogeneity and heterogeneity. Our approach is based on centered Dirichlet process mixtures (Yang and Dunson 2010), which allow uncertainty about the distributional forms. We show that assumptions about the joint distribution of the brand and time-specific constants and the residuals in the pricing equation can have a significant impact on the estimates of utility parameters and price elasticities.

Previous researchers have been either agnostic about the distributional forms for the unobserved variables, as in a GMM approach, or have assumed normally distributed unobserved variables (Villas-Boas and Winer 1999, Yang et al. 2003, Chintagunta et al. 2005). Assuming a parametric distribution leads to efficiency gains when the true distribution is used but may distort inferences otherwise. Villas-Boas (2007) and Park and Gupta (2009) point out that such an assumption of normality could be inconsistent with some marginal cost function specifications given a price-setting process, while being consistent with others. Methods based on the GMM are inherently more robust but can be less efficient than likelihood-based approaches. Here, we show how using a nonparametric Bayesian framework gives the benefits of robustness and enhanced efficiency when compared with parametric models with misspecified distributions.

Our nonparametric approach is related to that of Conley et al. (2008), who use Dirichlet process mixtures for instrumental variable estimation in linear models. We use centered Dirichlet process mixtures (CDPM) instead because identification restrictions are needed on the nonparametric distributions in the context of discrete choice models. We show how the CDPM can be used in the context of discrete choice models within a data-augmentation framework. Our approach for handling endogeneity can be considered a robust extension of the control-function method because it nonparametrically determines the appropriate control function to use in a given situation.

In addition, it allows a single-step estimation procedure without the need for additional procedures to calculate the uncertainty in parameter estimates.

We also use the CDPM approach to model heterogeneity in our framework. Heterogeneity is typically handled in discrete choice settings using latent class models, or via parametric distributions such as the multivariate normal or a finite mixture of normal distributions (Geweke and Keane 1999, 2001; Rossi et al. 2005). Researchers have also used the Dirichlet process (Ansari and Mela 2003, Ansari and Iyengar 2006, Kim et al. 2004) to accommodate discrete representations of heterogeneity in choice models. Recently, Burda et al. (2008) use Dirichlet process mixtures to nonparametrically specify a continuous distribution of heterogeneity. However, none of the above papers have simultaneously considered the endogeneity problem. Our CDPM approach to modeling heterogeneity can be regarded as a nonparametric extension of the finite mixture of the normals approach in that the CDPM uses a countably infinite mixture of normals, but it automatically infers the number of mixture components that are appropriate for a given data set while taking into account this additional source of uncertainty. Our approach is capable of flexibly accommodating situations that may be characterized by multimodality, skewness, outlying observations, and misspecification of functional form for the utilities without having to build specific models for each situation. In addition to the above benefits, Bayesian methods allow the incorporation of prior information, when available, and are an inherently small sample in nature. In contrast, the small sample properties of other estimation procedures such as the GMM are not well understood in such complex contexts.

By applying our methods to both simulated and scanner panel data sets, we show that distributional assumptions about endogeneity errors impact parameter inference and price elasticity estimates significantly. Our simulations show that the CDPM approach is capable of recovering the true parameters and price elasticities under many different assumptions for the endogeneity and heterogeneity distributions. Specifically, we show that when the true distribution is a normal, the CDPM is capable of mimicking the normal with some loss in efficiency compared with the true parametric model. In contrast, we find that a parametric model based on multivariate normal distributions does a poor job in recovering the parameter estimates and price elasticities when the errors come from nonnormal distributions that are multimodal, skewed, or heavy tailed. We apply our model to two scanner panel data sets involving household cleaner and shampoo categories. We find that parameter estimates are sensitive to the choice of distributional forms and that the CDPM yields evidence

of multimodality, skewness, and outlying observations. Model comparison statistics based on the widely applicable information criterion (WAIC; Watanabe 2010) and the deviance information criterion (DIC; Spiegelhalter et al. 2002) also indicate the superiority of the CDPM specification.

The rest of this paper is organized as follows. Section 2 introduces our modeling framework and presents the Dirichlet process and centered Dirichlet process mixtures. Section 3 describes our simulation and reports the results. Section 4 details the application and discusses the results obtained from applying alternative models on the two panel data sets. Section 5 concludes the paper with a discussion of its limitations and highlights areas of future research. All other details of the analysis are located in the appendices.

2. Model

In this section, we describe our semiparametric approach for handling endogeneity and heterogeneity in a discrete choice setting. As in the literature, we follow a random utility framework (Guadagni and Little 1983). We assume that on any given shopping trip (e.g., a store week), consumers choose either a single unit of the brand that gives the highest utility within a product category or an outside option (e.g., decide not to purchase within the category). Let J be the number of brands available in the category. The different choice alternatives can then be indexed by $j = 0, 1, \ldots, J$, where $j = 0$ refers to the "outside good," or the no-choice option. Let consumers be indexed by $i = 1, \ldots, I$. The choices made by consumer i are observed over $t = 1, \ldots T_i$ shopping trips. The utility u_{ijt} that consumer i receives from product j on trip t depends on the observed and unobserved attributes of the product and takes the following form:

$$
\begin{aligned}
u_{i0t} &= \epsilon_{i0t}, \quad j = 0, \\
u_{ijt} &= \mathbf{x}'_{jt}\boldsymbol{\beta}_i - \alpha_i p_{jt} + \eta_{jt} + \epsilon_{ijt}, \quad j = 1, \ldots, J.
\end{aligned} \tag{1}
$$

The vector $\mathbf{x}_{jt}$ contains nonprice marketing variables such as feature and display activities for brand j on trip t as well as brand dummies (i.e., brand-specific intercepts), and p_{jt} represents the price paid for the brand on trip t. The parameter vector $\boldsymbol{\beta}_i$ represents the consumer's response sensitivities to these marketing variables, and α_i captures the price sensitivity of the consumer. There are two types of unobserved variables (η_{jt} and ϵ_{ijt}) in the utility equation for a brand. The demand shock η_{jt} is common across all consumers who shop in a store in a given week and represents the average utility that these consumers obtain from the unobserved attributes of product j.

Such unobserved product attributes could include shelf space and shelf location in the store, or the presence of store coupons, for the week in which the trip is made, all of which are unobserved by the researcher. Because some of these unobserved factors could be common across brands, we allow the demand shocks to be correlated across the different brands in a store in a given week. The error ϵ_{ijt} represents factors that vary independently and identically distributed (i.i.d.) over brands, consumers, and purchase occasions. Assuming these are extreme value results in a logit model, whereas an assumption of normality yields a probit choice model.

The price for each product typically depends on all its attributes, both observed and unobserved. Thus, the prices in the utility equation are correlated with the demand shocks η_{jt}. Ignoring these unobserved attributes, therefore, can result in endogeneity bias and inconsistent parameter estimates. Previous researchers have handled this endogeneity problem using either a full information or a limited information approach (e.g., control functions as in Petrin and Train 2010). In the full information approach, the price-setting process for the firms is explicitly modeled using a game-theoretic framework and the actual prices in the data are assumed to be the equilibrium outcome of such a game (Yang et al. 2003, Villas-Boas and Zhao 2005). A number of different price-setting processes have been explored in the literature, including marginal cost pricing and Nash equilibrium pricing for single and multiproduct firms or retailers (Sudhir 2001). Such an explicit modeling of the price-setting process can yield efficiency gains if the correct process is used. However, it is unclear whether prices in the market place are indeed the outcome of an equilibrium because managers may not know enough about competition for the typical common knowledge assumptions to be correct. Moreover, even if the prices are from an equilibrium, the actual game is unobservable, and there is always a danger that the price-setting process is misspecified. In such cases, the wrong model of the supply side can potentially contaminate the demand side parameters (Berry 2003, Dubé and Chintagunta 2003). Another concern with such a structural approach is that it is often unclear whether the equilibrium of a particular game being assumed is unique, and this has implications for the use of the structural model to examine the effect of policy changes (Berry et al. 1995).

In contrast, the limited information approach (Villas-Boas and Winer 1999) is agnostic about the price-setting process and can therefore be considered more flexible and robust. In this paper, we follow such an approach and assume that the pricing equations for the J brands can be written as

$$
p_{jt} = \mathbf{z}'_{jt}\boldsymbol{\gamma}_j + \omega_{jt}, \quad j = 1, \ldots, J, \tag{2}
$$

where $\mathbf{z}_{jt}$ contains an intercept and the instrumental variables that are correlated with the price but are independent of the common demand shock, η_{jt}, in the utility specification. The error ω_{jt} in the pricing equation represents unobserved factors that affect costs. Endogeneity arises if η_{jt} and ω_{jt} are correlated. Given the possible presence of shared unobservables across brands and the equilibrium considerations that the price of one brand may depend on the demand shocks of all brands, we assume that the price shocks ω_{jt}'s and the demand shocks η_{jt}'s are all mutually correlated.

Rivers and Vuong (1988), Ching (2010), and Petrin and Train (2010) present alternative approaches to handling endogeneity that do not require the joint modeling of the demand and price shocks. Ching (2010) allows the demand shocks to directly enter the pricing equation in a linear fashion. Such a specification, however, amounts to assuming a normal joint distribution.[1] Petrin and Train (2010) suggest a two-step procedure in which residuals from the pricing equation are used in the demand model as a control function. This approach requires specifying the functional form for the control function. Our semiparametric approach toward joint modeling of these shocks can be considered a robust extension of the control-function method because it nonparametrically determines the appropriate control function to use in a given situation. Moreover, it offers advantages that stem from joint modeling of the errors when compared with a two-step control-function approach.

We show in this paper that assumptions about the joint distribution of $\boldsymbol{\eta}_t = \{\eta_{1t}, \ldots, \eta_{Jt}\}$, and $\boldsymbol{\omega}_t = \{\omega_{1t}, \ldots, \omega_{Jt}\}$ can impact inferences regarding the other parameters in the utility function. Previous researchers have either used approaches (e.g., based on GMM) that makes no specific assumptions about this joint distribution or have assumed a parametric distribution such as the normal (Villas-Boas and Winer 1999, Park and Gupta 2009, Yang et al. 2003). The GMM approach, although robust to misspecification, can be inefficient in the context of linear instrumental variable models, as has been shown by Conley et al. (2008). The assumption of joint normality, if true, can lead to efficient inference, but otherwise can distort conclusions. The normality assumption can be inconsistent with some price-setting behaviors of firms (Villas-Boas 2007). The presence of outlying observations or the misspecification of the utility function can also result in nonnormal errors. We therefore model this joint distribution flexibly using a Bayesian nonparametric approach. In particular, we assume that

the vector of unobserved variables $\boldsymbol{\zeta}_t = \{\boldsymbol{\eta}_t, \boldsymbol{\omega}_t\}$ is distributed according to a centered Dirichlet process mixture.

2.1. Centered Dirichlet Process Mixtures for Endogeneity

We assume that the $\boldsymbol{\zeta}_t$'s are independently drawn from an unknown continuous distribution that is centered at zero. Note that the systematic component of the utility includes brand-specific intercepts to capture the mean attraction of each choice alternative. These brand intercepts are substantively important in brand choice contexts (e.g., research involving brand equity), and researchers are also often interested in characterizing the heterogeneity in brand intercepts for purposes of preference segmentation. Hence, they cannot be treated as nuisance parameters that are integrated out of the analysis. Given their inclusion in the systematic part of the utility, the unobserved component $\boldsymbol{\zeta}_t$ needs to have an expectation of zero.

We model the distribution for $\boldsymbol{\zeta}_t$ as a mean mixture of normals $N(\boldsymbol{\nu}_t, \boldsymbol{\Omega})$, with the mixing distribution over the means $\boldsymbol{\nu}_t$ being an unknown distribution G, which is common for all brands. We let the prior for this mixing distribution be a centered Dirichlet process $CDP(G_0, \kappa)$, with concentration parameter κ and base distribution G_0. This gives the following hierarchy for the distribution of the $\boldsymbol{\zeta}_t$'s:

$$\boldsymbol{\zeta}_t \sim N(\boldsymbol{\nu}_t, \boldsymbol{\Omega}),$$
$$\boldsymbol{\nu}_t \sim G, \qquad (3)$$
$$G \sim CDP(G_0, \kappa).$$

The covariance matrix $\boldsymbol{\Omega}$ and the concentration parameter κ are given priors at a higher level. The centered Dirichlet process (Yang and Dunson 2010) is a generalization of the Dirichlet process (DP) introduced by Ferguson (1973). We now briefly review the basic properties of these processes.

2.1.1. Dirichlet Processes and Dirichlet Process Mixtures.
In Bayesian nonparametrics, the Dirichlet process is used to model the uncertainty about the functional form of an unknown distribution G and can thus be considered a distribution over distributions. A Dirichlet process prior for G is determined by two parameters: a base distribution function G_0 that sets the location of the Dirichlet process prior and a positive concentration parameter κ. Realizations from the Dirichlet process are discrete with probability one, which implies that the resulting $\boldsymbol{\nu}_t$ that draws from G will be grouped into clusters. The discrete nature of the DP can be made precise by looking at its constructive definition via the stick-breaking representation

---

[1] Ching (2010) suggests a general polynomial of demand shocks to approximate the pricing policy function but uses a linear specification because of data limitations.

because of Sethuraman (1994). According to this definition, $G \sim \mathrm{DP}(G_0, \kappa)$ implies that

$$G = \sum_{h=1}^{\infty} \pi_h \delta_{\theta_h},$$

$$\pi_h = V_h \prod_{l<h}(1 - V_l), \quad V_h \sim \mathrm{Beta}(1, \kappa), \; \theta_h \overset{iid}{\sim} G_0, \quad (4)$$

where δ_{θ_h} denotes a discrete measure concentrated at random atom θ_h, $h = 1, \ldots, \infty$; and V_h is an infinite sequence of stick-breaking probabilities (see Figure 1(a)). The construction of the probabilities π_h (i.e., the weights for the atoms) can be understood using the following metaphor of breaking a stick into successive segments. Starting with a stick of length one, the first break is at $\pi_1 = V_1$. The remaining portion $1 - V_1$ is then recursively broken to obtain $\pi_2 = (1 - V_1)V_2$, $\pi_3 = (1 - V_1)(1 - V_2)V_3$, and so on. Such a representation ensures that the infinite sum of the weights converges to one (see Figure 1(b)). The concentration parameter κ measures the strength of belief in the prior guess G_0. Larger values of κ generate smaller values of the stick-breaking probabilities, V_h, resulting in a sampled distribution G that mimics the base distribution G_0. In contrast, for small values of κ, the sampled G is likely to place most of its mass on a few atoms.

Inferences in DP models can use either a Polya-urn representation of the DP (Blackwell and MacQueen

Figure 1 **(a) Stick-Breaking Representation of Dirichlet Process; (b) Stick-Breaking Metaphor for the Probabilities of Dirichlet Process**

1973, Ansari and Mela 2003, Ansari and Iyengar 2006, Conley et al. 2008) or the stick-breaking one. We choose the latter because it allows a straightforward implementation of the centering restriction using a truncation approximation.

Because the realizations of the DP are discrete in nature, and the unobserved variables in the utility and pricing equations are best considered to be continuous, one can mix the DP with a continuous distribution such as the normal to yield a countably infinite mixture of normals. Any smooth density can then be accurately approximated using such a Dirichlet process mixture (DPM; Antoniak 1974). However, we need a continuous distribution that is centered at zero to model the variability in ζ_t. This cannot be ensured by merely choosing the base distribution G_0 to have a zero mean. In such a case, the prior expectation of the mean of G is zero, but the posterior expectation of the mean of G can differ significantly from this prior expectation and can bias inferences about the utility parameters and price elasticities. Hence, we use centered Dirichlet process mixtures to enforce this identification constraint.

2.1.2. Centered Dirichlet Process Mixtures. We follow Yang et al. (2010) and specify a CDPM distribution for the endogeneity errors as follows:[2]

$$\zeta_t \sim \mathrm{N}(\nu_t, \Omega),$$

$$\nu_t \sim G,$$

$$G = \sum_{h=1}^{\infty} V_h^E \prod_{l<h}(1 - V_l^E)\delta_{\theta_h}, \quad (5)$$

$$\theta_h = \theta_h^* - \psi_G^*,$$

$$\theta_h^* \sim G_0, \quad V_h^E \sim \mathrm{Beta}(1, \kappa^E),$$

where G_0 is a multivariate normal distribution with specified mean and variance. Our choice of a normal distribution for G_0 yields computational advantages because of conjugacy and allows us to nonparametrically deviate from the normal, which is a common choice for parametric models.

In the above, ψ_G^* is the unconstrained mean under the $\mathrm{DP}(G_0, \kappa)$ prior. The mean ψ_G^* and the variance Σ_G^* of the DP prior can be written as

$$\psi_G^* = \sum_{h=1}^{\infty} V_h^E \prod_{l<h}(1 - V_l^E)\theta_h^*,$$

$$\Sigma_G^* = \sum_{h=1}^{\infty} V_h^E \prod_{l<h}(1 - V_l^E)(\theta_h^* - \psi_G^*)(\theta_h^* - \psi_G^*)'. \quad (6)$$

The above construction implies that the induced mixture of normal distributions has a mean equal to

---

[2] In the rest of this paper, we use superscript $*$ to denote parameters from unconstrained Dirichlet processes.
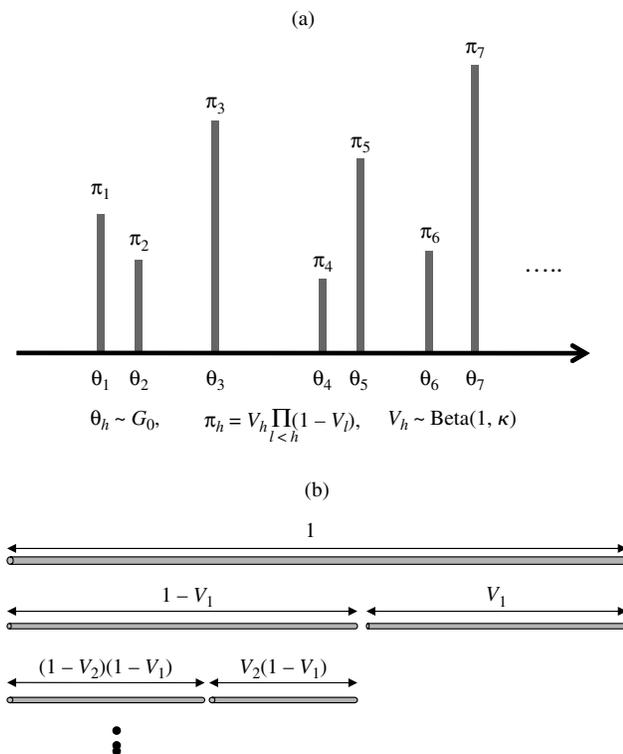
$E(\boldsymbol{\zeta}_t \mid G) = \mathbf{0}$ and a variance $V(\boldsymbol{\zeta}_t \mid G) = \boldsymbol{\Sigma}_G^* + \boldsymbol{\Omega}$, thus yielding a flexible prior that respects the identification constraint.

This completes our discussion of endogeneity priors. We now move to a discussion of consumer heterogeneity.

2.2. Centered Dirichlet Process Mixtures for Heterogeneity

Consumers are heterogeneous in their price sensitivities α_i and other response parameters $\boldsymbol{\beta}_i$. Incorporating such consumer heterogeneity is crucial for accurate parameter inference, substantive concerns, and targeting. There is a vast literature in marketing on different ways of modeling unobserved heterogeneity. Previous researchers have used finite mixtures (i.e., latent class), normal distributions, and finite mixtures of normals to model the variation in the individual-level parameters. Of these, the finite mixture of normals approach yields the greatest flexibility in modeling heterogeneity because it allows for multimodality, which may be important if different segments exist in the market place. However, it has the disadvantage that one needs to separately determine the number of components of the mixture. In contrast, when CDPM distributions are used, the number of components are automatically determined based on the data.

To model heterogeneity, we begin by rewriting the utility function in Equation (1). The utility function can be written as a sum of two components. The first component is a composite of factors that vary across products and time but do not vary across consumers, whereas the second component includes quantities that vary across consumers. Thus, we have

$$u_{ijt} = [\mathbf{x}'_{jt}\bar{\boldsymbol{\beta}} - \bar{\alpha}p_{jt} + \eta_{jt}] + \mathbf{w}'_{jt}\boldsymbol{\lambda}_i + \epsilon_{ijt}, \tag{7}$$

where $\mathbf{w}'_{jt} = (\mathbf{x}'_{jt}, -p_{jt})$. In the above equation, $\bar{\boldsymbol{\beta}}$ and $\bar{\alpha}$ are population means and represent the aggregate response tendencies in the population. Denoting the term in square brackets that does not vary across consumers as μ_{jt}, we can rewrite the utility function as

$$u_{ijt} = \mu_{jt} + \mathbf{w}'_{jt}\boldsymbol{\lambda}_i + \epsilon_{ijt}. \tag{8}$$

Because the variables in $\mathbf{w}_{jt}$ are also included as part of μ_{ij}, the individual coefficients $\boldsymbol{\lambda}_i$ are constrained to have a zero mean for identification. Analogous to our treatment of endogeneity, we flexibly model the heterogeneity using a CDPM of the following form:

$$\boldsymbol{\lambda}_i \sim \mathrm{N}(\boldsymbol{\varpi}_i, \boldsymbol{\Lambda}),$$

$$\boldsymbol{\varpi}_i \sim F,$$

$$F = \sum_{h=1}^{\infty} V_h^H \prod_{l<h}(1 - V_l^H)\delta_{\boldsymbol{\vartheta}_h}, \tag{9}$$

$$\boldsymbol{\vartheta}_h = \boldsymbol{\vartheta}_h^* - \boldsymbol{\psi}_F^*,$$

$$\boldsymbol{\vartheta}_h^* \sim F_0, \quad V_h^H \sim \mathrm{Beta}(1, \kappa^H),$$

where F_0 is a multivariate normal distribution with specified mean and variance. As before, $\boldsymbol{\psi}_F^*$ is the unconstrained mean of the corresponding Dirichlet process, and $\boldsymbol{\Lambda}$ is the covariance matrix that is estimated from the data.

2.3. Full Model

Bringing together the endogeneity and heterogeneity submodels, i.e., Equations (5) and (9), respectively, our modeling framework can be succinctly represented as follows:

$$u_{ijt} = \mu_{jt} + \mathbf{w}'_{jt}\boldsymbol{\lambda}_i + \epsilon_{ijt},$$

$$\boldsymbol{\mu}_t = \mathbf{x}_t\bar{\boldsymbol{\beta}} - \bar{\alpha}\mathbf{p}_t + \boldsymbol{\eta}_t,$$

$$\mathbf{p}_t = \mathbf{z}_t\boldsymbol{\gamma} + \boldsymbol{\omega}_t,$$

$$\boldsymbol{\zeta}_t \sim \mathrm{CDPM}(G_0, \kappa^E),$$

$$\boldsymbol{\lambda}_i \sim \mathrm{CDPM}(F_0, \kappa^H),$$

$$\epsilon_{ijt} \overset{\text{iid}}{\sim} \text{normal or extreme value.} \tag{10}$$

Estimation proceeds via Markov chain Monte Carlo (MCMC) methods based on data augmentation of the utility u_{ijt} and the mean utility μ_{jt} as well as a truncation approximation of the CDPM such that the resulting full conditional distributions are all available in closed form. The full conditional distributions used in our MCMC scheme are provided in Appendix A, §A.1.

Inferences about parameters and the number of components in Dirichlet process mixtures depend to some extent on the prior for κ and the choice of the base distribution G_0. In particular, the number of clusters could depend on the variance of the base measure. In practice, data are standardized before analysis and the base distribution is chosen to have close to unit variance. Unlike most applications in statistics, we use large data sets in our applications and thus inferences are mostly driven by the information contained in the likelihood. In the rest of this paper, we will employ the probit framework for the utilities. The logit can also be applied with appropriate modifications to our estimation scheme. We use the MCMC draws in tandem with the Geweke-Hajivassiliou-Keane (GHK) algorithm (Geweke 1989, Hajivassiliou and McFadden 1998, Keane 1994) to compute price elasticities and model comparison statistics WAIC and DIC. Details regarding these statistics are reported in Appendix A, §A.2.

3. Simulated Data

In this section, we apply simulated data to study the suitability of modeling the distributions of choice endogeneity and individual heterogeneity using CDPM. We demonstrate that the proposed model is able to properly recover the utility parameters and

price elasticities irrespective of the underlying distributions for the errors. We show that when the actual endogeneity and heterogeneity errors are normally distributed, our approach recovers preference parameters with some loss in efficiency. More importantly, we illustrate that when the normality assumption is violated and the data come from multimodal, skewed, or heavy-tailed distributions only models based on the CDPM are able to recover the true parameters and elasticities accurately. In such instances, models that assume normal distributions for endogeneity and heterogeneity errors yield incorrect estimates of the utility parameters and price elasticities. For the simulation study, we mainly focus on inferences when data come from multimodal distributions. However, we also report briefly, in the end of this section, on the performance of our model applied on data coming from unimodal, but skewed and heavy-tailed distributions. We now describe our simulated data.

3.1. Simulated Multimodal Data

We simulate four different data sets with varying distributional assumptions on the endogeneity and heterogeneity errors according to a 2×2 design involving two factors.

• The first design factor varies the distribution of choice endogeneity. That is, we simulate the endogeneity errors $\zeta_t = \{\eta_t, \omega_t\}$ from either a single multivariate normal distribution or from a mixture of seven multivariate normal distributions. In the latter case, we choose the weights, means, and covariance matrices of the component distributions to ensure a multimodal distribution that is centered at zero.

• The second design factor varies the heterogeneity distribution that characterizes the differences in the individual-level utility parameters. Analogous to the first factor, we assume that the person-specific coefficients, λ_i's, come from either a single multivariate normal distribution or from a mixture of seven multivariate normal distributions.

Note that we deliberately refrain from using a Dirichlet process mixture for generating the endogeneity and heterogeneity errors because our interest is in showing how the CDPM framework can flexibly handle a variety of data sets that are not directly based on it. The resulting data sets are as follows:

• D1: This uses normal endogeneity and heterogeneity components.

• D2: This uses the mixture distribution for endogeneity but the normal distribution for heterogeneity.

• D3: This uses normal endogeneity and the mixture distribution for heterogeneity.

• D4: This uses mixture distributions for both the endogeneity errors and heterogeneity coefficients.

We use the same sample size for all four simulated data sets with $I = 300$ individuals, $T = 400$ trips, and two alternative brands $J = 2$ plus an outside good on each choice occasion. We assume that the vector x_t contains the two brand constants and a single brand-specific exogenous variable (e.g., promotion). In addition, we assume a single brand-specific endogenous variable (e.g., price). The "prices" are generated according to Equation (2). The exogenous variable and the "instruments" z_t in the price equations are assumed to be uniformly distributed independently across brands and observations. The brand-specific intercepts and the price and promotion coefficients are allowed to be heterogeneous across individuals. The true values for all utility and distributional parameters for generating the four data sets are reported in Appendix B.

3.2. Models

We estimate four models on each of the simulated data sets. The models are as follows:

• M1: In this model, we use multivariate normal distributions for both endogeneity and heterogeneity errors. This serves as a benchmark model, which our other models can be compared with to assess the gains from a semiparametric approach.

• M2: In this model, we use a multivariate normal distribution to capture heterogeneity but allow the endogeneity errors to follow a CDPM. A comparison of M2 with M1 can reveal the benefits of a nonparametric specification for the endogeneity distribution.

• M3: Here, we specify the endogeneity distribution to be multivariate normal but assume that the heterogeneity distribution is a CDPM.

• M4: In the last model, we assume two different CDPM specifications for the endogeneity and heterogeneity errors. Thus, M4 is the full model that we consider in this paper.

Details about the priors for the four models can be found in Appendix A, §A.1. We now discuss the estimation results for each of the four data sets.

3.3. Results

The results for all 16 model runs are based on 50,000 MCMC draws after discarding an initial set of 50,000 burn-in draws. Convergence is assessed by monitoring the trace plots of the model parameters.

3.3.1. D1: Normal Endogeneity and Normal Heterogeneity.
Table 1 reports the posterior mean of the utility parameters, the price elasticities, and their posterior standard deviations for the four models when applied to D1. In addition, the table also contains the estimates of the concentration parameters of the endogeneity and heterogeneity CDPM distributions for models M2, M3, and M4. Comparing the posterior means across models, we can conclude that all four models recover well the true parameters and elasticities. However, the standard deviations for the utility parameters in models that involve the CDPM are somewhat higher than those from M1. This loss

Table 1   MCMC Estimation Results for D1

| Parameter	Truth	M1 Est.	M1 Std.	M2 Est.	M2 Std.	M3 Est.	M3 Std.	M4 Est.	M4 Std.
Intercept 1	1.3	1.325	0.060	1.328	0.179	1.326	0.102	1.328	0.177
Intercept 2	0.8	0.839	0.061	0.856	0.182	0.863	0.103	0.857	0.184
Promotion	0.7	0.628	0.031	0.652	0.036	0.653	0.058	0.650	0.050
Price	1.1	1.100	0.032	1.110	0.041	1.108	0.057	1.110	0.042
κ^E	—	—	—	22.211	1.965	—	—	22.237	1.974
κ^H	—	—	—	—	—	17.905	1.698	17.903	1.730
e11	−1.130	−1.131	0.037	−1.132	0.035	−1.130	0.043	−1.133	0.037
e12	0.616	0.621	0.022	0.622	0.023	0.621	0.029	0.621	0.024
e21	0.365	0.372	0.011	0.372	0.011	0.371	0.013	0.372	0.012
e22	−1.549	−1.557	0.045	−1.558	0.045	−1.556	0.055	−1.557	0.048
WAIC	—	2.827		3.151		2.830		3.293	
DIC	—	2.804		3.290		2.807		3.341	

in efficiency is not surprising given that the actual data come from normal distributions. However, the differences in standard deviations are small enough, particularly for the elasticities, that the qualitative conclusions based on the models remain unaffected. Finally, we see that both the WAIC and the DIC statistics are the lowest for model M1 and thus point toward the correct model.

We can also assess how well the full model M4 recovers the unimodal nature of the heterogeneity and endogeneity distributions. As an example, Figure 2 shows the actual distribution as well as the posterior predictive distribution for endogeneity errors associated with the price equations for brands 1 and 2. It is clear from the figure that the CDPM specification is able to recover the unimodal nature of the true distribution. The recovery is similar for the other endogeneity components as well as for the multivariate distribution of the heterogeneity errors.

3.3.2. D2: Mixture Endogeneity and Normal Heterogeneity. Table 2 reports the posterior mean and standard deviations of the parameters and the

elasticities for the four models on D2. Recall that, in this data set, the actual endogeneity errors follow a mixture of seven normal distributions. In M1 and M3, the endogeneity component assumes a normal distribution and is therefore misspecified. We see that the posterior means for the parameters in these models differ significantly from the true values. This difference is particularly pronounced for the brand intercepts. In contrast, M2 and M4 yield estimates that are close to the truth. This is because the CDPM specification for the endogeneity component in these models is able to mimic the true finite mixture of normals distribution. It is also interesting to note that model M3 is unable to recover the true parameters despite its flexible handling of heterogeneity using the CDPM. Thus, flexibility in the heterogeneity component is not sufficient to account for misspecification in the endogeneity component. We can see from the table that the posterior standard deviations, particularly for the endogenous price variable, are also inflated for models M1 and M3 when compared with those for M2 and M4. Focusing on the elasticities, we see that M2 and M4 are not only better at recovering the

Figure 2   Actual Distribution of the Endogeneity Errors of D1 (Left) and the Posterior Predictive Distribution of M4 (Right)
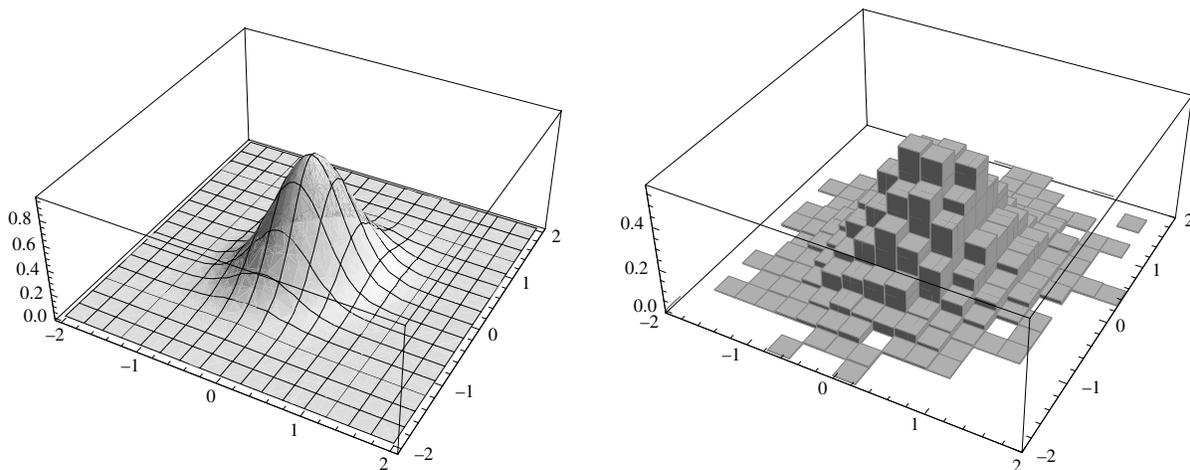
Table 2   MCMC Estimation Results for D2

Parameter	Truth	M1 Est.	M1 Std.	M2 Est.	M2 Std.	M3 Est.	M3 Std.	M4 Est.	M4 Std.
Intercept 1	1.3	1.543	0.291	1.206	0.248	1.674	0.326	1.426	0.286
Intercept 2	0.8	1.826	0.349	0.818	0.244	1.986	0.398	0.802	0.282
Promotion	0.7	0.626	0.148	0.768	0.048	0.623	0.152	0.768	0.053
Price	1.1	1.174	0.146	1.076	0.046	1.273	0.188	1.026	0.051
κ^E	—	—	—	6.217	1.215	—	—	8.661	1.495
κ^H	—	—	—	—	—	2.240	0.704	2.512	0.757
e11	−0.273	−0.292	0.037	−0.258	0.012	−0.317	0.047	−0.256	0.013
e12	0.102	0.111	0.014	0.099	0.005	0.120	0.018	0.098	0.006
e21	0.083	0.089	0.009	0.079	0.003	0.095	0.011	0.079	0.004
e22	−0.369	−0.393	0.043	−0.350	0.014	−0.420	0.054	−0.347	0.016
WAIC	—	2.557		2.108		2.304		2.342	
DIC	—	2.439		2.219		2.321		2.237	

elasticities but do so efficiently, given the smaller standard deviations. We also find that the WAIC and DIC statistics are the lowest for the correct model (M2). These results clearly indicate that flexibility in modeling endogeneity is important to avoid the possibility of misleading results.

3.3.3. D3: Normal Endogeneity and Mixture Heterogeneity. Table 3 reports the results for the four models on D3. We find that all four models do a reasonably good job in recovering the true utility parameters and price elasticities. In addition, we can see that the posterior standard deviations are also comparable across the model. Recall that, in this data set, the heterogeneity coefficients come from a mixture of normal distributions, whereas the endogeneity errors are distributed normal. The good performance of models M1 and M2 (both models use a parametric multivariate normal distribution for the heterogeneity component) indicates that inferences regarding the utility parameters are not overly sensitive to the misspecification of the heterogeneity distribution.

An examination of the recovery of the endogeneity and heterogeneity distributions (figures are not included for brevity) reveals that models M3 and M4

recover the exact shape of the seven-normal mixture distribution for the heterogeneity, whereas model M1 is unable to do so. Thus, to the extent that uncovering the pattern of individual differences is important for managerial actions (e.g., segmentation), it is still preferable to use the semiparametric specification.

3.3.4. D4: Mixture Endogeneity and Mixture Heterogeneity. Table 4 reports the results for the four models on D4. It is clear from the table that models M1 and M3, which assume a parametric distribution for the endogeneity errors, do a poor job of recovering the true parameters as well as the price elasticities. Moreover, the associated posterior standard deviations for these models are inflated compared with those from models M2 and M4. The latter two models yield estimates and elasticities that are close to their true values because both accommodate endogeneity in a flexible fashion. We also see that the WAIC and DIC identify the right model M4.

Finally, as before, the full model is able to recover multimodality of the underlying true distributions. Figure 3 shows a bivariate marginal distribution of the endogeneity errors (for the demand shock of brand 1 and the price shock of brand 2) from the

Table 3   MCMC Estimation Results for D3

Parameter	Truth	M1 Est.	M1 Std.	M2 Est.	M2 Std.	M3 Est.	M3 Std.	M4 Est.	M4 Std.
Intercept 1	1.3	1.098	0.123	1.251	0.213	1.284	0.130	1.214	0.115
Intercept 2	0.8	0.744	0.109	0.738	0.204	0.743	0.121	0.770	0.113
Promotion	0.7	0.604	0.034	0.603	0.035	0.606	0.075	0.587	0.067
Price	1.1	1.059	0.038	1.052	0.038	1.060	0.050	1.059	0.051
κ^E	—	—	—	5.651	0.930	—	—	1.967	0.592
κ^H	—	—	—	—	—	4.670	1.039	6.006	1.295
e11	−0.602	−0.584	0.019	−0.584	0.019	−0.584	0.020	−0.585	0.022
e12	0.294	0.290	0.011	0.290	0.011	0.288	0.011	0.291	0.012
e21	0.184	0.182	0.007	0.182	0.007	0.181	0.007	0.182	0.007
e22	−0.982	−0.976	0.030	−0.975	0.030	−0.974	0.031	−0.975	0.034
WAIC	—	2.455		2.650		2.406		2.614	
DIC	—	2.423		2.483		2.422		2.382	

Table 4   MCMC Estimation Results for D4

Parameter	Truth	M1 Est.	M1 Std.	M2 Est.	M2 Std.	M3 Est.	M3 Std.	M4 Est.	M4 Std.
Intercept 1	1.3	2.087	0.459	1.358	0.339	2.307	0.649	1.348	0.274
Intercept 2	0.8	1.678	0.546	0.680	0.323	1.871	0.787	0.783	0.267
Promotion	0.7	0.582	0.201	0.716	0.043	0.569	0.208	0.717	0.059
Price	1.1	1.419	0.250	1.065	0.038	1.519	0.371	1.082	0.042
κ^E	—	—	—	12.908	1.888	—	—	8.358	1.504
κ^H	—	—	—	—	—	17.513	3.405	6.073	1.316
e11	−0.230	−0.311	0.057	−0.225	0.009	−0.331	0.083	−0.228	0.010
e12	0.125	0.174	0.035	0.121	0.006	0.185	0.051	0.122	0.007
e21	0.102	0.144	0.028	0.100	0.005	0.153	0.039	0.101	0.006
e22	−0.608	−0.835	0.157	−0.596	0.024	−0.888	0.227	−0.599	0.027
WAIC	—	2.448		2.360		2.445		2.253	
DIC	—	2.404		2.284		2.217		2.188	

actual seven-normal mixture (top left), the posterior predictive distribution of M4 (top right), and the posterior predictive distribution of M1 (bottom). It is clear from the figure that the posterior predictive distribution of M4 mimics the shape of the true distribution

remarkably well. However, the posterior predictive distribution of M1 differs from the actual distribution. Figure 4 shows the univariate marginals from the true four-dimensional heterogeneity distribution and from the posterior predictive distribution of M4. Once

Figure 3   Bivariate Marginals of the Endogeneity Errors (for the Demand Shock of Brand 1 and the Price Shock of Brand 2) from the Actual Distribution of D4 (Top Left), the Posterior Predictive Distribution of M4 (Top Right), and the Posterior Predictive Distribution of M1 (Bottom)

Figure 4   Marginals of the Heterogeneity Errors from the Actual Distribution of D4 (Left) and the Posterior Predictive Distribution of M4 (Right)
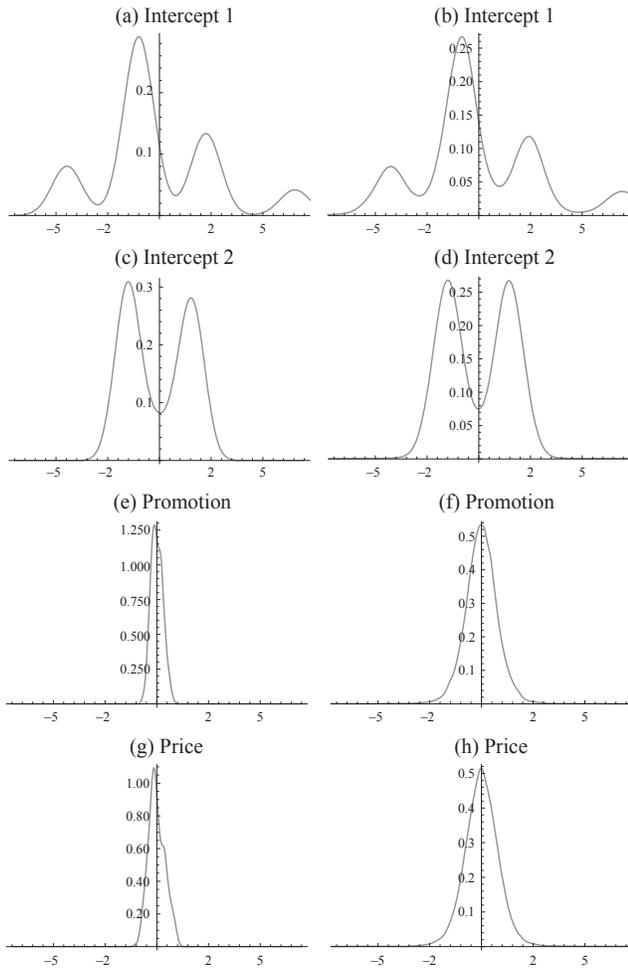

(a) Intercept 1

(b) Intercept 1

(c) Intercept 2

(d) Intercept 2

(e) Promotion

(f) Promotion

(g) Price

(h) Price

Table 5   MCMC Estimation Results for the Skewed and Heavy-Tailed Data

Parameter	Truth	M1		M2	
		Est.	Std.	Est.	Std.
Intercept	1.3	0.878	0.393	1.084	0.475
Promotion	0.8	0.719	0.173	0.739	0.114
Price	1.1	0.838	0.210	1.015	0.169
κ^E	—	—	—	2.379	0.674
e11	−0.602	−0.465	0.114	−0.556	0.090
WAIC	—	1.958		1.510	
DIC	—	1.711		1.484	

it is evident that the model comparison statistics are capable of pointing toward the right model.[3]

So far, we have focused on model performance when data come from multimodal distributions. We now briefly discuss how the different models perform when the endogeneity errors come from skewed and heavy-tailed distributions.
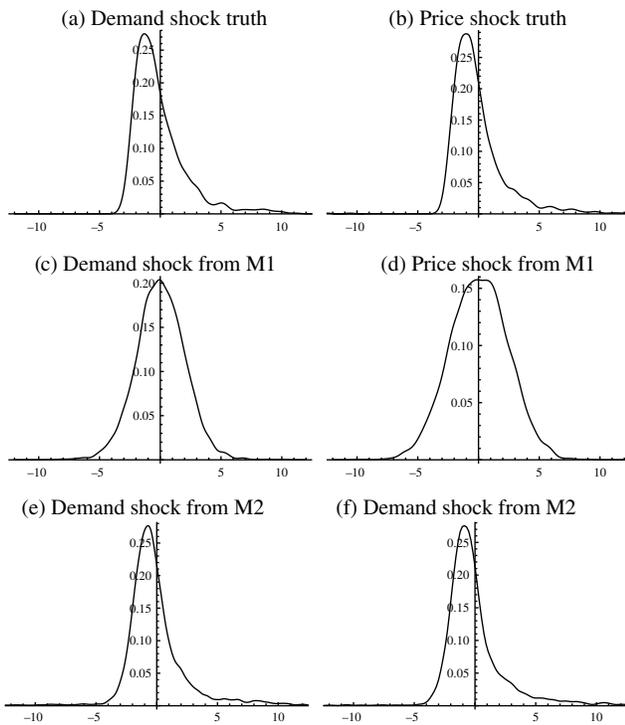
3.4.   Simulated Skewed and Heavy-Tailed Data

Here, we show the superiority of the proposed CDPM endogeneity model on skewed and heavy-tailed data. In this exercise, for brevity, we only consider a data set involving the choice between a single brand and an outside option. To generate skewness and heavy tails in the data, we use a noncentral t-distribution (Fisher 1931) and a Gaussian copula (Song 2000). Specifically, we simulate the demand shock, η_t, via a noncentral t distribution with degrees of freedom $= 3$, and a noncentrality parameter $= 6$. Next we generate the price shock, ω_t, from another noncentral t distribution with degrees of freedom $= 4$ and a noncentrality parameter $= 7$. The joint distribution of η_t and ω_t is then constructed using a bivariate Gaussian copula with correlation $= 0.7$ on the two marginal distributions. For simplicity, the distribution of heterogeneity errors is kept as a multivariate normal. All other characteristics of the data are similar to the previously constructed data sets. Given that the heterogeneity errors are normal, we estimate only M1 and M2 on this data. Table 5 shows the parameter and elasticity estimates. Figure 5 compares the posterior predictive distributions of the endogeneity errors from M1 and M2.

It is clear from Table 5 that M2 outperforms M1 in recovering the model parameters and the price elasticity. Also, we see from Figure 5 that M2 does a better job (compared with M1) of mimicking the shape of the

again, we can see that the posterior predictive distribution of M4 mimics the actual heterogeneity distribution well.

3.3.5.   Summary of Results. Summarizing the results across the four data sets, it is evident that the estimates of utility parameters and price elasticities are sensitive to the choice of distributions for the endogeneity and heterogeneity errors. In particular, the functional form of the endogeneity distribution appears to more noticeably impact the results. It is also clear from the above that the full model does a good job of recovering these parameters and elasticities. The CDPM specification is able to produce more efficient estimates in most situations, and it causes some loss in efficiency only when the data purely come from a normal distribution. Moreover, the full model properly reconstructs the shape of the underlying distributions, whether they are unimodal or multimodal in nature, thereby giving an accurate portrayal of the segmentation structure of the market. Lastly,

---

[3] We also investigate the importance of centering for Dirichlet process mixtures by comparing the parameter estimates from variants of model M4 that use unconstrained DPM for the endogeneity and/or heterogeneity distributions. We find that the unconstrained versions result in a lack of identification, bad mixing, and poor recovery of model parameters. The results and discussion are provided in Appendix C.

Figure 5 Marginals of the Endogeneity Errors from the Actual Distribution of the Skewed and Heavy-Tailed Data (Top), the Posterior Predictive Distribution of M1 (Middle), and the Posterior Predictive Distribution of M2 (Bottom)

(a) Demand shock truth

(b) Price shock truth

(c) Demand shock from M1

(d) Price shock from M1

(e) Demand shock from M2

(f) Demand shock from M2

distributions of demand and price shocks. The model comparison statistics again identify the right model.

4. Real Data Applications

In this section, we apply our method to two real data sets from the recently available consumer IRI marketing data set (Bronnenberg et al. 2008). There are 50 IRI markets included in the original data set. Store-level data are available for all 50 markets. However, household panel data are available for only two of these markets (i.e., Eau Claire, WI and Pittsfield, MA). The panelists' shopping behavior is observed from January 2001 to December 2006. We chose two categories–household cleaner and shampoo—to illustrate our approach. The descriptive statistics of our samples are presented in Table 6. In selecting the households to include in our samples, we first restrict the panel to households who made at least one grocery trip in each of the six years. Then for each product category, we restrict the sample to households that are in the top 95% of the distribution of purchase frequencies within the category. The second criterion results in differences in sample size across the two categories.

To construct brand choice observations from the original data sets, we first match the panel choices with the store sales using the latter to build choice sets at brand level. We include all top market share brands that together account for at least 70% of the

Table 6 Descriptive Statistics of the Household Cleaner and Shampoo Data Sets

Category	No. of observations	No. of households	No. of brands	Price mean	Promotion mean	Share of outside option (%)
Household cleaner	8,415	360	8	1.29	0.10	20
Shampoo	18,469	753	7	3.50	0.30	32

sales within a category and group all other small share brands (i.e., < 4% each) into a single brand as the outside option for that category. The price variable in Table 6 refers to the regular list price per unit size that is provided in the original data. In both categories, this corresponds to price for a 16 oz. pack. As is common in brand choice studies, we obtain brand price as sales-weighted mean of the prices of the universal product codes (UPCs) belonging to that brand. We acknowledge, however, that such aggregation across UPCs can have implications for the magnitude of own- and cross-price elasticities that are inferred from the data (Song 2010). The promotion variable in the table refers to the promotion probability for a brand in a store week. This variable is obtained by setting promotion equal to 1 for a 5% or higher temporary price reduction for a UPC and then computing a sales-weighted average of this variable across the UPCs of the brand.

We specify the utility of each brand as in Equation (1). Utility is assumed to be a function of the brand intercept, the price, and an exogenous promotion variable. Endogeneity concerns arise for the price variable p_{jt} because it is potentially correlated with the unobserved demand shock η_{jt}. For example, omitted brand attributes, such as shelf-space allocation, may affect consumers' choice process, but researchers do not observe these variables in the data. Also, prices may be set strategically based on their potential impact on demand; hence, the brand prices are not exogenous. Instruments are therefore needed to separate the exogenous variation from the endogenous variation in price. We follow (Hausman 1996, Nevo 2001) and exploit the multimarket nature of the data sets to select valid instruments. Assuming that market-specific price variations are independent across markets, prices in other geographic markets are valid instruments for prices in any given market because prices of the same brand are correlated because of common production costs. We therefore use the prices of the same brand in the same week in other markets that are geographically separated from a given market as instruments.[4] Given the current

---

[4] Details regarding the identities of the markets that are used in each case are available from the authors upon request. The R squares in the regression of prices on the instruments are 0.56 for household cleaner and 0.88 for shampoo.

Table 7   Estimation Results of the Household Cleaner Category

Parameter	M1 Est.	M1 Std.	M2 Est.	M2 Std.	M3 Est.	M3 Std.	M4 Est.	M4 Std.
Intercept 1	2.561	0.346	1.894	0.373	2.541	0.337	1.770	0.324
Intercept 2	1.569	0.273	0.991	0.293	1.614	0.261	0.856	0.245
Intercept 3	1.034	0.285	0.394	0.297	1.137	0.264	0.426	0.255
Intercept 4	1.389	0.321	0.659	0.358	1.410	0.296	0.646	0.285
Intercept 5	1.185	0.306	0.292	0.353	1.107	0.314	0.201	0.305
Intercept 6	0.164	0.286	−0.645	0.331	0.160	0.307	−0.855	0.289
Intercept 7	0.695	0.259	−0.036	0.298	0.875	0.247	0.122	0.260
Intercept 8	0.204	0.224	−0.415	0.239	0.289	0.250	−0.571	0.218
Price	0.414	0.058	0.311	0.056	0.395	0.053	0.288	0.055
Promotion	0.146	0.020	0.109	0.025	0.129	0.021	0.117	0.024
κ^E	—	—	39.634	3.417	—	—	67.530	5.753
No. of components endo.	—	—	13.800	0.000	—	—	14.000	0.000
κ^H	—	—	—	—	10.573	1.626	7.814	1.614
No. of components het.	—	—	—	—	8.000	0.000	3.000	0.000
WAIC	10.597		10.456		10.688		10.413	
DIC	8.298		8.884		8.486		8.143	

Table 8   Estimation Results of the Shampoo Category

Parameter	M1 Est.	M1 Std.	M2 Est.	M2 Std.	M3 Est.	M3 Std.	M4 Est.	M4 Std.
Intercept 1	0.382	0.096	0.404	0.104	0.372	0.106	0.468	0.113
Intercept 2	0.605	0.287	0.603	0.287	0.520	0.266	0.767	0.303
Intercept 3	−0.997	0.099	−1.028	0.113	−1.006	0.108	−0.959	0.122
Intercept 4	0.358	0.244	0.375	0.246	0.345	0.238	0.521	0.274
Intercept 5	0.505	0.274	0.479	0.281	0.437	0.275	0.618	0.321
Intercept 6	0.029	0.306	0.033	0.300	−0.017	0.281	0.164	0.360
Intercept 7	−1.517	0.108	−1.572	0.129	−1.539	0.122	−1.510	0.133
Price	0.849	0.103	0.900	0.107	0.848	0.102	0.956	0.118
Promotion	0.266	0.015	0.272	0.017	0.253	0.016	0.257	0.017
κ^E	—	—	16.280	1.798	—	—	77.153	7.819
No. of components endo.	—	—	21.275	0.447	—	—	43.000	0.000
κ^H	—	—	—	—	17.132	1.917	8.826	1.315
No. of components het.	—	—	—	—	17.000	0.000	7.000	0.000
WAIC	14.842		15.116		14.751		13.756	
DIC	9.583		9.453		10.519		9.237	

data, however, we cannot rule out the possibility that common demand shocks, such as nationwide marketing campaigns, may exist across locations, thereby affecting the validity of the chosen instruments.

4.1. Results

We estimate the four models M1 to M4 on the data sets. These model specifications mirror those for the simulation. Parameter estimates and the model comparison statistics for the four models in each of the two product categories are based on 50,000 MCMC draws after discarding an initial set of 100,000 burn-in draws.

Tables 7 and 8 present the posterior means and the associated posterior standard deviations for the two product categories under each of the four model specifications. One can see from the results that there are considerable differences in the parameter estimates between the parametric model M1 and the full model M4. The brand intercepts are severely impacted by the distributional assumptions, particularly within the

household cleaner category. The price coefficient also exhibits greater variation across the four models in household cleaner than in shampoo. We find considerable variations in the price elasticity estimates as well,[5] thus reinforcing the importance of the nonparametric specification in modeling endogeneity.

Figure 6 shows the posterior predictive distributions of the endogeneity errors for the eight brands in the household cleaner category. It is clear from the figure that the endogeneity errors exhibit varying degrees of multimodality, skewness, and outlying observations. Skewness and thick tails are particularly pronounced in the demand shocks, whereas multimodality is evident in the price shocks. We also observe a moderate level of nonnormality in the

---

[5] We report the own-price elasticity estimates in Appendix D for both categories. The cross-price elasticity results are available from the authors upon request.

Figure 6   Marginals of the Posterior Predictive Distribution of the Endogeneity Errors in the Household Cleaner Category
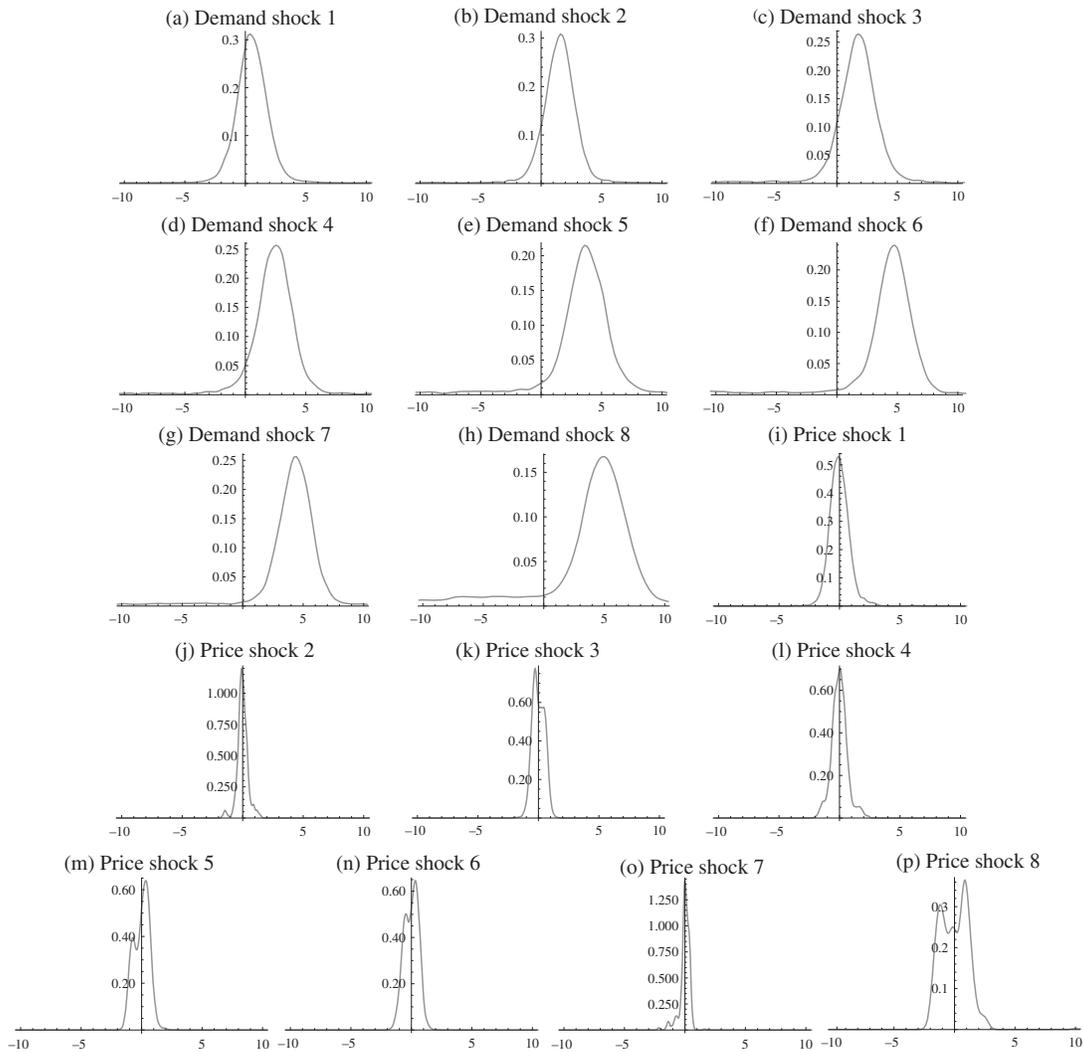
posterior predictive distributions of heterogeneity errors in the two categories. The WAIC and DIC point toward the full model M4 in both product categories, indicating that it is important to accommodate the above patterns in the demand and price shocks.

In the current model setup, price endogeneity is generated via the correlations between demand and price shocks. We can recover such correlations using the variance matrix $V(\zeta_t \mid G) = \Sigma_G^* + \Omega$, where Σ_G^* is obtained from (6). Note that we allow every demand shock to be correlated with each other and with all price shocks; therefore, we end up with a 16-by-16 error variance matrix for the household cleaner data and a 14-by-14 error variance matrix for the shampoo data. In the interest of brevity, in Table 9, we show only the within-brand correlations between demand and price shocks. (The estimates of the full correlation matrices can be obtained from the authors upon request.) One can see from the table that there are moderate and statistically significant correlations

between the demand and price shocks, justifying the endogeneity concerns in both categories.

The results from the two real data sets indicate that parameter estimates and price elasticities are sensitive to the choice of distributional forms for endogeneity and heterogeneity errors. The results also

Table 9   Within-Brand Correlations Between Demand and Price Shocks

	Household cleaner		Shampoo	
	Est.	Std.	Est.	Std.
Brand 1	−0.013	0.072	0.136	0.040
Brand 2	−0.179	0.050	0.106	0.037
Brand 3	−0.066	0.063	0.097	0.036
Brand 4	0.060	0.057	0.009	0.096
Brand 5	−0.159	0.066	0.026	0.070
Brand 6	0.119	0.044	0.103	0.033
Brand 7	0.016	0.055	0.029	0.034
Brand 8	0.213	0.061	—	—

suggest greater gains from modeling endogeneity in a flexible manner. Given the uncertainty about the appropriate form to use in a particular application, the CDPM approach allows us to infer nonparametrically such distributions based on the data. The extent of the endogeneity effect (i.e., the correlations between demand and price shocks) can also be obtained automatically from the MCMC scheme as reported in Table 9.

5. Conclusion

The growing interest in accommodating endogenous regressors in discrete choice settings has resulted in a proliferation of methodological approaches. These methods differ in their robustness and in the efficiency of the parameter estimates they produce. In this paper we show how misspecification of the endogeneity and heterogeneity distributions in discrete choice models can impact the recovery of model parameters and elasticity estimates. This has consequences for managerial actions because these are typically based on price elasticity estimates. Considering that the true distributions of random shocks are unknown in any given situation, we advocate the use of a flexible and robust approach for inferring such distributions via a semiparametric Bayesian framework. Our approach relies on centered Dirichlet process mixtures to accommodate this uncertainty about distributional forms. This offers a flexible and robust alternative that does well under a variety of situations, particularly when data come from multimodal, skewed, or heavy-tailed distributions, without having to specifically build separate models for these different situations. It is important to note that the semiparametric approach achieves such robustness without impacting the efficiency of parameter estimates significantly. Instead, we find that the CDPM can yield more efficient estimates than a parametric model with distributional forms that are misspecified. Moreover, our approach also offers all the traditional advantages of Bayesian procedures, such as the ability to incorporate prior information and generate exact inference in small samples, in tandem with the benefits of numerical stability arising from the use of MCMC methods.

We apply our framework to both simulated and scanner panel data sets. The study on the simulated data offers the benefit of knowing the true data generating process, and thus we are able to investigate to what extent true utility parameters and elasticities are recovered by our approach. Our results clearly indicate that the estimates for the population mean, which capture the overall response tendencies in the population, are impacted significantly by the assumptions for the endogeneity distributions. We find that a model that relies on parametric distributions such

as the normal does a poor job of recovering the aggregate utility parameters as well as the elasticities when the true underlying distributions are nonnormal. In contrast, the CDPM based models are able to recover well the parameters irrespective of the true underlying distribution for the data. The CDPM does well in both multimodal and skewed/heavy-tailed settings, whereas parametric models distort inferences about the mean as well as the statistical significance of the parameters in such settings.

We also find that the CDPM performs well in recovering the shape of the underlying distributions and thereby yields proper qualitative conclusions about the nature of variability in the population. It is interesting to note that the CDPM is able to recover well unimodal distributions (as in the case of normal errors). Moreover, in such cases, the loss in efficiency that arises from the necessity to accommodate the uncertainty about distributional forms in the CDPM approach is not large and mostly happens on brand intercepts. Our simulation results show that the parameter and elasticity estimates exhibit greater sensitivity to the misspecification of the endogeneity distribution than that of the heterogeneity distribution, and thus it is crucial to model the endogeneity errors in a flexible fashion.

In our real data applications, we show that the parameter and elasticity estimates differ across model specifications and that the model comparison statistics support the full model. The posterior predictive distributions based on the MCMC runs also indicate the possibilities of multimodality, skewness, and outlying observations in the data, thus supporting the need for a semiparametric alternative. Summarizing across our applications, we can conclude that our Bayesian semiparametric approach offers a robust alternative to models that rely on the normal distribution for handling both the endogeneity and heterogeneity errors. In addition, it is clear from the results that our model yields more accurate elasticity estimates. To the extent that managerial action is often closely related to estimates of price elasticities, our approach is superior in guiding managerial decision making.

Our enquiry can be extended in multiple directions. We use centered Dirichlet process priors as the underlying building blocks. The literature on Bayesian nonparametrics is rapidly evolving and offers other alternatives. Future researchers could use probit stick-breaking processes (Rodríguez and Dunson 2011) or normalized inverse-Gaussian priors (Lijoi et al. 2005) instead of Dirichlet processes. In this paper, we assume that the utility errors are normally distributed to yield a probit model. Our methods can be modified via a data augmentation scheme for extreme value distributed utilities to handle the multinomial logit case. We focus on the consequences of misspecified distributional forms. It would

be interesting to investigate in future research the robustness of the CDPM in recovering model parameters under different equilibrium assumptions of the supply side (e.g., Bertrand-Nash). Moreover, although we primarily study the impact of distributional forms on inferences, the assumption on the price shock being additive in Equation (2) could be inconsistent with some marginal cost specification, given a price-setting process. We leave a detailed exploration of this issue for future research. Finally, the current analysis is mainly applied to individual-level choice models. Researchers can further explore whether our approach can be used for aggregate data situations. It is unclear, however, whether aggregate market share data contain enough information to identify nonparametric distributions.

Supplemental Material

Acknowledgments

Appendix A. MCMC

A.1. Full Conditionals
We now list the full conditionals for the parameters in model M4 along with their prior distributions. To implement the stick-breaking process in MCMC, we follow the truncation approximation as in Ishwaran and Zarepour (2000) and Ishwaran and James (2001), and we set the truncation points at $N^E = N^H = 150$ for both endogeneity and heterogeneity CDPMs. The wide truncation ensures a close approximation to the Dirichlet processes.

(1) $\boldsymbol{\mu}_t$: The mean utility for the constrained Dirichlet process follows the two equations

$$\begin{aligned}
\boldsymbol{\mu}_t &= \mathbf{X}_t \bar{\boldsymbol{\beta}} - \bar{\alpha} \mathbf{p}_t + \boldsymbol{\eta}_t, \\
\mathbf{p}_t &= \mathbf{Z}_t \boldsymbol{\gamma} + \boldsymbol{\omega}_t.
\end{aligned} \tag{A1}$$

Because we assume $\boldsymbol{\eta}_t$ and $\boldsymbol{\omega}_t$ are distributed as a joint normal, $N(\boldsymbol{\nu}_t, \boldsymbol{\Omega})$, the prior for $\boldsymbol{\mu}_t$ is a marginal distribution, $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, conditional on $\boldsymbol{\omega}_t$, where $\boldsymbol{\mu}_0 = \mathbf{X}_t \bar{\boldsymbol{\beta}} + (\boldsymbol{\Omega}_{\eta\omega} \boldsymbol{\Sigma}_w^{-1} - \bar{\alpha} \mathbf{I}) \mathbf{p}_t + \boldsymbol{\nu}_{\eta t} - \boldsymbol{\Omega}_{\eta\omega} \boldsymbol{\Sigma}_w^{-1} (\mathbf{Z}_t \boldsymbol{\gamma} + \boldsymbol{\nu}_{\omega t})$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Omega}_\eta - \boldsymbol{\Omega}_{\eta\omega} \boldsymbol{\Sigma}_w^{-1} \boldsymbol{\Omega}_{\eta\omega}'$. The model for $\boldsymbol{\mu}_t$ is

$$\mathbf{u}_{it} - \mathbf{w}_t \boldsymbol{\lambda}_i = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_{it}, \tag{A2}$$

so the posterior distribution of $\boldsymbol{\mu}_t$ is a multivariate normal distribution with precision matrix, $\boldsymbol{\Sigma}_0^{-1} + I_t \mathbf{I}$, and mean vector, $(\boldsymbol{\Sigma}_0^{-1} + I_t \mathbf{I})^{-1} (\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + \sum_{i=1}^{I_t} (\mathbf{u}_{it} - \mathbf{w}_t \boldsymbol{\lambda}_i))$, where I_t is the number of individuals who made purchases at t.

(2) $\boldsymbol{\mu}_t^*$: The mean utility for the unconstrained Dirichlet process differs from $\boldsymbol{\mu}_{jt}$ by a constant:

$$\boldsymbol{\mu}_t^* = \boldsymbol{\mu}_t + (\boldsymbol{\Sigma}_0^{-1} + I_t \mathbf{I})^{-1} \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\psi}_{G1}^* - \boldsymbol{\Omega}_{\eta\omega} \boldsymbol{\Sigma}_w^{-1} \boldsymbol{\psi}_{G2}^*). \tag{A3}$$

(3) $\bar{\boldsymbol{\beta}}$, $\bar{\alpha}$, and $\boldsymbol{\gamma}_j$: Let $\mathbf{b}$ and $\mathbf{X}_t$ denote the parameters and the data at the right-hand side of (A1), respectively, and

write $\mathbf{Y}_t = (\boldsymbol{\mu}_t, \mathbf{p}_t)$. Then (A1) is transformed into

$$\mathbf{Y}_t - \boldsymbol{\nu}_t = \mathbf{X}_t \mathbf{b} + \mathbf{e}_t, \tag{A4}$$

where $\mathbf{e}_t \sim N(\mathbf{0}, \boldsymbol{\Omega})$. For conjugacy, we assume the prior distribution for $\mathbf{b}$ is $N(\mathbf{b}_0, \boldsymbol{\Sigma}_b)$. Define $T = \max\{T_i\}$, and the posterior distribution for $\mathbf{b}$ is a multivariate normal distribution $N(\hat{\mathbf{b}}, \hat{\boldsymbol{\Sigma}}_b)$, where

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_b^{-1} &= \boldsymbol{\Sigma}_b^{-1} + \sum_{t=1}^T \mathbf{X}_t' \boldsymbol{\Omega}^{-1} \mathbf{X}_t, \\
\hat{\mathbf{b}} &= \hat{\boldsymbol{\Sigma}}_b \left(\boldsymbol{\Sigma}_b^{-1} \mathbf{b}_0 + \sum_{t=1}^T \mathbf{X}_t' \boldsymbol{\Omega}^{-1} (\mathbf{Y}_t - \boldsymbol{\nu}_t) \right).
\end{aligned} \tag{A5}$$

(4) $\boldsymbol{\Omega}$: We assume the prior distribution is an inverse Wishart distribution, $\boldsymbol{\Omega} \sim IW(\rho_\Omega, \mathbf{R}_\Omega)$. From (A4), $\mathbf{e}_t = (\mathbf{Y}_t - \boldsymbol{\nu}_t - \mathbf{X}_t \mathbf{b}) \sim N(\mathbf{0}, \boldsymbol{\Omega})$, the posterior distribution is also an inverse Wishart:

$$IW\left(\rho_\Omega + T, R_\Omega + \sum_{t=1}^T (\mathbf{Y}_t - \boldsymbol{\nu}_t - \mathbf{X}_t \mathbf{b})(\mathbf{Y}_t - \boldsymbol{\nu}_t - \mathbf{X}_t \mathbf{b})' \right). \tag{A6}$$

(5) $\boldsymbol{\lambda}_i$: The constrained individual coefficient has a prior distribution $\boldsymbol{\lambda}_i \sim N(\boldsymbol{\varpi}_i, \boldsymbol{\Lambda})$. From (8), $(\mathbf{u}_{it} - \boldsymbol{\mu}_t) = \mathbf{w}_t \boldsymbol{\lambda}_i + \boldsymbol{\epsilon}_{it}$, the posterior distribution is thus a multivariate normal distribution, $N(\hat{\boldsymbol{\varpi}}_i, \hat{\boldsymbol{\Lambda}}_i)$, where

$$\begin{aligned}
\hat{\boldsymbol{\Lambda}}_i^{-1} &= \boldsymbol{\Lambda}^{-1} + \sum_{t=1}^{T_i} \mathbf{w}_t' \mathbf{w}_t, \\
\hat{\boldsymbol{\varpi}}_i &= \hat{\boldsymbol{\Lambda}}_i \left(\boldsymbol{\Lambda}^{-1} \boldsymbol{\varpi}_i + \sum_{t=1}^{T_i} \mathbf{w}_t' (\mathbf{u}_{it} - \boldsymbol{\mu}_t) \right).
\end{aligned} \tag{A7}$$

(6) $\boldsymbol{\lambda}_i^*$: The unconstrained individual coefficients differs from $\boldsymbol{\lambda}_i$ by a constant, i.e., $\boldsymbol{\lambda}_i^* = \boldsymbol{\lambda}_i + \hat{\boldsymbol{\Lambda}}_i \boldsymbol{\Lambda}^{-1} \boldsymbol{\psi}_F^*$.

(7) $\boldsymbol{\Lambda}$: We assume the prior distribution is an inverse Wishart, $IW(\rho_\Lambda, \mathbf{R}_\Lambda)$. Because the model is $\boldsymbol{\lambda}_i \sim N(\boldsymbol{\varpi}_i, \boldsymbol{\Lambda})$, the posterior distribution for $\boldsymbol{\Lambda}$ is

$$IW\left(\rho_\Lambda + I, R_\Lambda + \sum_{i=1}^I (\boldsymbol{\lambda}_i - \boldsymbol{\varpi}_i)(\boldsymbol{\lambda}_i - \boldsymbol{\varpi}_i)' \right). \tag{A8}$$

(8) u_{ijt}: According to (1), u_{ijt} has a prior distribution, $u_{ijt} \sim N((\mu_{jt} + \mathbf{w}_{jt}' \boldsymbol{\lambda}_i), 1)$, for $j = 1, \ldots, J$, and $u_{i0t} \sim N(0, 1)$, for $j = 0$. An alternative k is chosen if and only if $u_{ikt} = \max_j \{u_{ijt}\}$, for $j = 0, 1, \ldots, J$. Therefore, the posterior for utility u_{ijt} is a truncated normal distribution with mean and variance being the same as the prior, and the truncation ensures the highest utility corresponds to the observed choice.

(9) $\boldsymbol{\theta}_h^*$: The location parameter $\boldsymbol{\theta}_h^*$ of the unconstrained Dirichlet process for endogeneity errors is drawn from the baseline distribution G_0, which we assume to be $N(\mathbf{0}, \mathbf{I})$. Let $\{\tilde{K}_1, \ldots, \tilde{K}_m\}$ denote the set of current m unique values of all location indicators $\{K_t^E\}_{\forall t}$, and write $\mathbf{Y}_t^* = (\boldsymbol{\mu}_t^*, \mathbf{p}_t)$. For each location that is not informed by the data, $h \in \{1, \ldots, N^E\} - \{\tilde{K}_1, \ldots, \tilde{K}_m\}$, the posterior is the same as the prior distribution, i.e., $\boldsymbol{\theta}_h^* \sim G_0$. But for each $h \in \{\tilde{K}_1, \ldots, \tilde{K}_m\}$, the data provide information through the model, $\boldsymbol{\nu}_t^* = (\mathbf{Y}_t^* - \mathbf{X}_t \mathbf{b}) \sim N(\boldsymbol{\theta}_h^*, \boldsymbol{\Omega})$, $\forall t$, s.t. $K_t^E = \tilde{K}_k$, where $k = 1, \ldots, m$. In this case, the posterior is a multivariate normal distribution,

$N(\hat{\boldsymbol{\theta}}_h, \hat{\boldsymbol{\Sigma}}_{\theta_h})$, where

$$\hat{\boldsymbol{\Sigma}}_{\theta_h}^{-1} = \mathbf{I} + \sum_{\forall t: K_t^E = \check{K}_h} \boldsymbol{\Omega}^{-1},$$

$$\hat{\boldsymbol{\theta}}_h = \hat{\boldsymbol{\Sigma}}_{\theta_h} \left(\sum_{\forall t: K_t^E = \check{K}_h} \boldsymbol{\Omega}^{-1} (\mathbf{Y}_t^* - \mathbf{X}_t \mathbf{b}) \right). \quad (A9)$$

(10) $\boldsymbol{\vartheta}_h^*$: The location parameter $\boldsymbol{\vartheta}_h^*$ of the unconstrained Dirichlet process for heterogeneity errors is drawn from the baseline distribution F_0, which is assumed to be $N(\mathbf{0}, \mathbf{I})$. Let $\{\check{K}_1, \ldots, \check{K}_n\}$ denote the set of current n unique values of all location indicators $\{K_i^H\}_{i=1}^I$. For each location that is not informed by the data, $h \in \{1, \ldots, N^H\} - \{\check{K}_1, \ldots, \check{K}_n\}$, the posterior is the same as the prior distribution, i.e., $\boldsymbol{\vartheta}_h^* \sim F_0$. But for each $h \in \{\check{K}_1, \ldots, \check{K}_n\}$, the data provide information through the model, $\boldsymbol{\lambda}_i^* \sim N(\boldsymbol{\vartheta}_h^*, \boldsymbol{\Lambda})$, $\forall i$, s.t. $K_i^H = \check{K}_k$, where $k = 1, \ldots, n$. In this case, the posterior is a multivariate normal distribution, $N(\hat{\boldsymbol{\vartheta}}_h, \hat{\boldsymbol{\Sigma}}_{\vartheta_h})$, where

$$\hat{\boldsymbol{\Sigma}}_{\vartheta_h}^{-1} = \mathbf{I} + \sum_{\forall i: K_i^H = \check{K}_h} \boldsymbol{\Lambda}^{-1}, \qquad \hat{\boldsymbol{\vartheta}}_h = \hat{\boldsymbol{\Sigma}}_{\vartheta_h} \left(\sum_{\forall i: K_i^H = \check{K}_h} \boldsymbol{\Lambda}^{-1} \boldsymbol{\lambda}_i^* \right). \quad (A10)$$

(11) K_t^E: The location indicator assigns each $\boldsymbol{\nu}_t^*$ to an appropriate location h in the Dirichlet process for endogeneity errors. It follows a discrete distribution, $K_t^E \sim \sum_{h=1}^{N^E} \pi_{ht} \delta_h$, where $\pi_{ht} \propto \pi_h^E \cdot N(\mathbf{Y}_t^* - \mathbf{X}_t \mathbf{b} \mid \boldsymbol{\theta}_h^*, \boldsymbol{\Omega})$. π_h^E is the prior probability assigned to location h and is given by

$$\pi_1^E = V_1^E \quad \text{and} \quad \pi_h^E = V_h^E \prod_{l < h} (1 - V_l^E)$$

$$\text{for } h = 2, \ldots, N^E. \quad (A11)$$

(12) K_i^H: The location indicator assigns each $\boldsymbol{\varpi}_i^*$ to an appropriate location h in the Dirichlet process for heterogeneity errors. It follows a discrete distribution, $K_i^H \sim \sum_{h=1}^{N^H} \pi_{hi} \delta_h$, where $\pi_{hi} \propto \pi_h^H \cdot N(\boldsymbol{\lambda}_i^* \mid \boldsymbol{\vartheta}_h^*, \boldsymbol{\Sigma})$. π_h^H is the prior probability assigned to location h and is given by

$$\pi_1^H = V_1^H \quad \text{and} \quad \pi_h^H = V_h^H \prod_{l < h} (1 - V_l^H)$$

$$\text{for } h = 2, \ldots, N^H. \quad (A12)$$

(13) V_h^E: The prior for the stick-breaking probability for endogeneity errors is $V_h^E \sim \text{Beta}(1, \kappa^E)$. Let M_h^E be the number of K_t^E's that point to location h, $\forall t$. According to Ishwaran and Zarepour (2000), the posterior distributions for these stick-breaking probabilities are

$$V_{N^E}^E = 1 \quad \text{and} \quad V_h^E \sim \text{Beta}\left(1 + M_h^E, \kappa^E + \sum_{l=h+1}^{N^E} M_l^E\right)$$

$$\text{for } h = 1, \ldots, N^E - 1. \quad (A13)$$

(14) V_h^H: The prior for the stick-breaking probability for heterogeneity errors is $V_h^H \sim \text{Beta}(1, \kappa^H)$. Let M_h^H be the number of K_i^H's that point to location h, $\forall i$. Then the posterior distributions are

$$V_{N^H}^H = 1 \quad \text{and} \quad V_h^H \sim \text{Beta}\left(1 + M_h^H, \kappa^H + \sum_{l=h+1}^{N^H} M_l^H\right)$$

$$\text{for } h = 1, \ldots, N^H - 1. \quad (A14)$$

(15) κ^E: A conjugate prior is a gamma distribution, i.e., $\kappa^E \sim \Gamma(d_1^E, d_2^E)$. Then the posterior is $\kappa^E \sim \Gamma(N^E + d_1^E - 1, d_2^E - \log(\pi_{N^E}^E))$.

(16) κ^H: The prior is $\kappa^H \sim \Gamma(d_1^H, d_2^H)$, and the posterior is $\kappa^H \sim \Gamma(N^H + d_1^H - 1, d_2^H - \log(\pi_{N^H}^H))$.

(17) $\boldsymbol{\nu}_t$: According to the location indicator K_t^E,

$$\boldsymbol{\nu}_t = \boldsymbol{\theta}_{K_t}^* - \boldsymbol{\psi}_G^*. \quad (A15)$$

(18) $\boldsymbol{\varpi}_i$: According to the location indicator K_i^H,

$$\boldsymbol{\varpi}_i = \boldsymbol{\vartheta}_{K_i}^* - \boldsymbol{\psi}_F^*. \quad (A16)$$

A.2. WAIC and DIC

Denote the entire set of unknown parameters as $\mathcal{W}$ and each of the $\mathcal{N}$ choice observations as $\mathcal{C}_k$, $k = 1, \ldots, \mathcal{N}$. In every MCMC iteration, we apply the GHK algorithm (Geweke 1989, Hajivassiliou and McFadden 1998, Keane 1994) to compute the likelihood, $\mathbb{P}(\mathcal{C}_k \mid \mathcal{W})$, for each observed choice and its associated price vector because the utilities and prices are a jointly distributed multivariate normal. Let $\mathbb{E}_{\mathcal{W}}[\cdot]$ be the expected value with respect to the posterior distribution of $\mathcal{W}$, i.e., averaging across MCMC iterations. According to Watanabe (2010), the widely applicable information criterion (WAIC) is defined as

$$\text{WAIC} = \frac{1}{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} \{ \mathbb{E}_{\mathcal{W}}[(\log \mathbb{P}(\mathcal{C}_k \mid \mathcal{W}))^2] - \mathbb{E}_{\mathcal{W}}[\log \mathbb{P}(\mathcal{C}_k \mid \mathcal{W})]^2$$
$$- \log \mathbb{E}_{\mathcal{W}}[\mathbb{P}(\mathcal{C}_k \mid \mathcal{W})] \}. \quad (A17)$$

We also compare alternative models using deviance information criterion (DIC, Spiegelhalter et al. 2002), which is written by

$$\text{DIC} = \frac{1}{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} \{ 2 \log \mathbb{P}(\mathcal{C}_k \mid \mathbb{E}_{\mathcal{W}}[\mathcal{W}]) - 2 \mathbb{E}_{\mathcal{W}}[\log \mathbb{P}(\mathcal{C}_k \mid \mathcal{W})]$$
$$- \log \mathbb{E}_{\mathcal{W}}[\mathbb{P}(\mathcal{C}_k \mid \mathcal{W})] \}. \quad (A18)$$

Appendix B. The Synthetic Data-Generating Process

Here we report the parameter values we use to simulate the data sets in §3. For D1, the variance matrices of the endogeneity errors and heterogeneity errors are, respectively,

$$\boldsymbol{\Omega} = \begin{bmatrix} 0.150 & 0.050 & 0.050 & 0.050 \\ & 0.200 & 0.050 & 0.050 \\ & & 0.150 & 0.050 \\ & & & 0.200 \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.232 & -0.258 & -0.183 & -0.304 \\ & 1.744 & 0.762 & 1.246 \\ & & 0.838 & 0.960 \\ & & & 1.672 \end{bmatrix}.$$

The mixture distribution for the endogeneity errors in D2 and D4 are mixed from seven multivariate normal components with weights = $\{0.2, 0.1, 0.2, 0.1, 0.1, 0.15, 0.15\}$. The mean vectors of these normal components are

$$\left\{ \begin{pmatrix} 1.77 \\ 4.60 \\ 0.13 \\ 4.92 \end{pmatrix}, \begin{pmatrix} 2.83 \\ 2.83 \\ 2.61 \\ -1.08 \end{pmatrix}, \begin{pmatrix} 4.60 \\ 1.77 \\ -3.57 \\ 3.39 \end{pmatrix}, \begin{pmatrix} 0.00 \\ -2.83 \\ -1.08 \\ -2.61 \end{pmatrix}, \right.$$

$$\left. \begin{pmatrix} -2.83 \\ 0.00 \\ -1.53 \\ 3.70 \end{pmatrix}, \begin{pmatrix} -6.36 \\ -2.12 \\ -0.48 \\ -6.69 \end{pmatrix}, \begin{pmatrix} -2.12 \\ -6.36 \\ 5.07 \\ -4.40 \end{pmatrix} \right\},$$

and the variance matrices are

$$\begin{bmatrix} 0.188 & 0.063 & 0.063 & 0.063 \\ & 0.250 & 0.063 & 0.063 \\ & & 0.188 & 0.063 \\ & & & 0.250 \end{bmatrix}, \quad \text{for } j = 2, 4, 6;$$

$$\begin{bmatrix} 0.250 & 0.088 & 0.088 & 0.088 \\ & 0.250 & 0.088 & 0.088 \\ & & 0.250 & 0.088 \\ & & & 0.250 \end{bmatrix}, \quad \text{for } j = 1, 3, 5, 7.$$

The mixture distribution for the heterogeneity errors in D3 and D4 are mixed from seven multivariate normal components with weights = {0.2, 0.1, 0.2, 0.325, 0.325, 0.1, 0.1}. The mean vectors of these normal components are

$$\left\{ \begin{pmatrix} 2.25 \\ 1.5 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 6.5 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2.25 \\ -1.5 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -1.5 \\ 0 \\ 0 \end{pmatrix}, \right.$$
$$\left. \begin{pmatrix} -1 \\ 1.5 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -4.5 \\ 1.5 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -4.5 \\ -1.5 \\ 0 \\ 0 \end{pmatrix} \right\}. \quad \text{(B1)}$$

We use the same variance matrix as the above Σ across the seven components.

Appendix C. Unconstrained DPM Models

We continue the discussion of CDPM by highlighting succinctly the importance of imposing the centering constraint on the Dirichlet process mixtures. Table C.1 reports parameter estimates for four models on data set D4. The first model is M4 and the other three are variational forms of M4 obtained by using different combinations of unconstrained Dirichlet process mixtures (instead of constrained) on the endogeneity and heterogeneity errors.

As can be seen from the table, models based on the unconstrained Dirichlet process mixture result in poor recovery of the parameter estimates because of identification problems. In particular, we see that the brand intercepts are impacted adversely when the endogeneity distribution

is unconstrained. We also see that all parameters are influenced because of using an unconstrained heterogeneity distribution. It is interesting to note, however, that the elasticity estimates remain relatively unaffected by this identification issue. In addition, the constrained specification is preferable because it offers additional advantages. For instance, it has better numerical properties because there is less chance of overflow and underflow errors and it is easier to assess convergence when one works with identifiable parameters.

Appendix D. Own-Price Elasticity Estimates for Real Data Applications

Here we present the own-price elasticities for the household cleaner category in Table D.1 and for the shampoo category in Table D.2. It is clear from the tables that models M1 and M3 report different elasticities when compared with those obtained from models M2 and M4. This again supports our view that the nonparametric specification is crucial in modeling endogeneity.

Table D.1   Own-Price Elasticity Estimates for the Household Cleaner Category

Parameter	M1 Est.	M1 Std.	M2 Est.	M2 Std.	M3 Est.	M3 Std.	M4 Est.	M4 Std.
e11	−1.519	0.214	−1.163	0.208	−1.453	0.223	−1.039	0.216
e22	−1.363	0.205	−1.012	0.207	−1.322	0.201	−0.939	0.203
e33	−1.512	0.236	−1.130	0.226	−1.423	0.206	−1.060	0.235
e44	−1.405	0.234	−1.079	0.208	−1.338	0.200	−0.981	0.229
e55	−2.005	0.363	−1.490	0.333	−1.912	0.361	−1.422	0.353
e66	−1.617	0.301	−1.219	0.294	−1.563	0.300	−1.086	0.298
e77	−1.584	0.278	−1.231	0.233	−1.521	0.241	−1.109	0.276
e88	−2.013	0.334	−1.522	0.267	−1.918	0.333	−1.395	0.334

Table D.2   Own-Price Elasticity Estimates for the Shampoo Category

Parameter	M1 Est.	M1 Std.	M2 Est.	M2 Std.	M3 Est.	M3 Std.	M4 Est.	M4 Std.
e11	−0.418	0.053	−0.462	0.062	−0.434	0.064	−0.479	0.065
e22	−1.145	0.223	−1.339	0.247	−1.195	0.260	−1.437	0.254
e33	−0.507	0.078	−0.547	0.087	−0.516	0.089	−0.577	0.089
e44	−1.448	0.206	−1.521	0.232	−1.360	0.243	−1.661	0.243
e55	−1.537	0.222	−1.628	0.251	−1.453	0.255	−1.776	0.256
e66	−1.511	0.294	−1.618	0.327	−1.586	0.339	−1.817	0.317
e77	−0.601	0.095	−0.639	0.105	−0.565	0.104	−0.680	0.108

Table C.1   Parameter Estimates from Unconstrained DPM Models

	Truth	M4 (constrained) Est.	M4 (constrained) Std.	M4 (uncons. het.) Est.	M4 (uncons. het.) Std.	M4 (uncons. endo.) Est.	M4 (uncons. endo.) Std.	M4 (uncons. both) Est.	M4 (uncons. both) Std.
Intercept 1	1.3	1.348	0.274	1.225	0.392	−3.061	1.593	4.233	1.197
Intercept 2	0.8	0.783	0.267	−0.028	0.384	−8.243	1.931	−6.720	1.364
Promotion	0.7	0.717	0.059	2.510	0.282	0.706	0.060	0.219	0.191
Price	1.1	1.082	0.042	1.172	0.041	1.056	0.039	0.999	0.058
e11	−0.230	−0.228	0.010	−0.227	0.010	−0.226	0.010	−0.227	0.009
e12	0.125	0.122	0.007	0.123	0.006	0.121	0.006	0.123	0.006
e21	0.102	0.101	0.006	0.101	0.005	0.100	0.005	0.101	0.005
e22	−0.608	−0.599	0.027	−0.601	0.026	−0.597	0.026	−0.602	0.024

References

Ansari A, Iyengar R (2006) Semiparametric Thurstonian models for recurrent choices: A Bayesian analysis. *Psychometrika* 71:631–657.

Ansari A, Mela C (2003) E-customization. *J. Marketing Res.* 40: 131–145.

Antoniak C (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* 2:1152–1174.

Berry S (2003) Comment: Bayesian analysis of simultaneous demand and supply by Yang, Chen and Allenby. *Quant. Marketing Econom.* 1:285–291.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63:841–890.

Blackwell D, MacQueen JB (1973) Ferguson distributions via Polya Urn schemes. *Ann. Statist.* 2:353–355.

Bronnenberg BJ, Kruger MW, Mela CF (2008) The IRI marketing data set. *Marketing Sci.* 27:745–748.

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

Ching AT (2010) Consumer learning and heterogeneity: Dynamics of demand for prescription drugs after patent expiration. *Internat. J. Indust. Organ.* 28:619–638.

Chintagunta P (2001) Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Sci.* 20:442–456.

Chintagunta P, Dubé JP, Goh KY (2005) Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models. *Management Sci.* 51:832–849.

Conley TG, Hansen CB, McCulloch RE, Rossi PE (2008) A semiparametric Bayesian approach to the instrumental variable problem. *J. Econometrics* 144:276–305.

Dubé JP, Chintagunta P (2003) Comment: Bayesian analysis of simultaneous demand and supply by Yang, Chen and Allenby. *Quant. Marketing Econom.* 1:293–298.

Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. *Ann. Statist.* 1:209–230.

Fisher RA (1931) Properties of Hh functions. *British Association Mathematical Tables*, Vol. 1 (Cambridge University Press, Cambridge, UK).

Geweke J (1989) Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57:1317–1339.

Geweke J, Keane MP (1999) Mixture of normals probit models. Hsiao C, Lahiri K, Lee LF, Pesaran MH, eds. *Analysis of Panels and Limited Dependent Variable Models* (Cambridge University Press, Cambridge, UK), 49–78.

Geweke J, Keane MP (2001) Computationally intensive methods for integration in econometrics. Heckman JJ, Leamer E, eds. *Handbook of Econometrics* (Elsevier Science, New York), 3463–3568.

Goolsbee A, Petrin A (2004) The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica* 72:351–381.

Guadagni PM, Little JDC (1983) A logit model of brand choice calibrated on scanner data. *Marketing Sci.* 3:203–238.

Hajivassiliou V, McFadden D (1998) The method of simulated scores for the estimation of LDV models. *Econometrica* 66:863–896.

Hausman JA (1996) Valuation of new goods under perfect and imperfect competition. Bresnahan TF, Gordon RJ, eds. *The Economics of New Goods*, Studies in Income and Wealth, Vol. 58 (University of Chicago Press, Chicago), 209–247.

Ishwaran H, James LF (2001) Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* 96:161–173.

Ishwaran H, Zarepour M (2000) Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* 87:371–390.

Keane MP (1994) A computationally practical simulation estimator for panel data. *Econometrica* 62:95–116.

Kim JG, Menzefricke U, Feinberg FM (2004) Assessing heterogeneity in discrete choice models using a Dirichlet process prior. *Rev. Marketing Sci.* 2:1–39.

Kuksov D, Villas-Boas JM (2008) Endogeneity and individual consumer choice. *J. Marketing Res.* 45:702–714.

Lijoi A, Mena RH, Prunster I (2005) Hierarchical mixture modeling with normalized inverse-gaussian priors. *J. Amer. Statist. Assoc.* 100:1278–1291.

Nevo A (2001) Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69:307–342.

Park S, Gupta S (2009) Simulated maximum likelihood estimator for the random coefficient logit model using aggregate data. *J. Marketing Res.* 46:531–542.

Petrin A, Train K (2010) A control function approach to endogeneity in consumer choice models. *J. Marketing Res.* 47:3–13.

Rivers D, Vuong QH (1988) Limited information estimators and exogeneity tests for simultaneous probit models. *J. Econometrics* 39:347–66.

Rodríguez A, Dunson DB (2011) Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Anal.* 6:145–178.

Rossi PE, Allenby GM, McCulloch R (2005) *Bayesian Statistics and Marketing*, 1st ed. (John Wiley & Sons, New York).

Sethuraman J (1994) A constructive definition of Dirichlet priors. *Statistica Sinica* 4:639–650.

Song I (2010) Item aggregates and price elasticity. *Seoul J. Bus.* 16:45–63.

Song PX-K (2000) Multivariate dispersion models generated from Gaussian copula. *Scandinavian J. Statist.* 27:305–320.

Spiegelhalter DJ, Best NG, Carlin BP, Linde A (2002) Bayesian measures of model complexity and fit. *J. Royal Statist. Soc., Series B* 64:583–639.

Sudhir K (2001) Structural analysis of manufacturer pricing in the presence of a strategic retailer. *Marketing Sci.* 20:244–264.

Villas-Boas JM (2007) A note on limited versus full information estimation in non-linear models. Working paper, University of California, Berkeley, Berkeley.

Villas-Boas JM, Winer RS (1994) Endogeneity in brand choice models. Mimeo, University of California, Berkeley, Berkeley.

Villas-Boas JM, Winer RS (1999) Endogeneity in brand choice models. *Management Sci.* 45:1324–1338.

Villas-Boas JM, Zhao Y (2005) Retailer, manufacturers, and individual consumers: Modeling the supply side in the ketchup marketplace. *J. Marketing Res.* 42:83–95.

Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Machine Learn. Res.* 11:3571–3594.

Yang M, Dunson DB (2010) Bayesian semiparametric structural equation models with latent variables. *Psychometrika* 75:675–693.

Yang M, Dunson DB, Baird D (2010) Semiparametric Bayes hierarchical models with mean and variance constraints. *Comput. Statist. Data Anal.* 54:2172–2186.

Yang S, Chen Y, Allenby GM (2003) Bayesian analysis of simultaneous demand and supply. *Quant. Marketing Econom.* 1:251–275.