Marketing Science

Exact MCMC for Choices from Menus—Measuring Substitution and Complementarity Among Menu Items

Tetyana Kosyakova, Thomas Otter, Sanjog Misra, Christian Neuerburg

Please scroll down for article—it is on subsequent pages

Exact MCMC for Choices from Menus—Measuring Substitution and Complementarity Among Menu Items

Tetyana Kosyakova,[a] Thomas Otter,[b] Sanjog Misra,[c] Christian Neuerburg[d]

[a] Frankfurt School of Finance & Management, 60322 Frankfurt am Main, Germany; [b] Goethe University Frankfurt, 60323 Frankfurt, Germany; [c] University of Chicago, Chicago, Illinois 60637; [d] Adidas, 91074 Herzogenaurach, Germany
Contact: t.kosyakova@fs.de, http://orcid.org/0000-0002-5450-5616 (TK); otter@marketing.uni-frankfurt.de (TO); sanjog.misra@chicagobooth.edu (SM); christian.neuerburg@adidas.com (CN)

Abstract. Choice environments in practice are often more complicated than the well-studied case of choice between perfect substitutes. Consumers choosing from menus or configuring products face choice sets that consist of substitutes, complements, and independent items, and the utility-maximizing choice corresponds to a particular item combination out of a potentially huge number of possible combinations. This reality is mirrored in menu-based choice experiments. The inferential challenge posed by data from such choices is in the calibration of utility functions that accommodate a mix of substitutes, complements, and independent items. We develop a model that not only accounts for heterogeneity in preferences, but also in what consumers perceive to be substitutes and complements and show how to perform Bayesian inference for this model based on the exact likelihood, despite its practically intractable normalizing constant. We characterize the model from first principles and show how it structurally improves on the multivariate probit model and on models that include cross-price effects in the utility function. We find empirical support for our model in a menu-based discrete choice experiment investigating demand for game consoles and accessories. Finally, we illustrate substantial implications from modeling substitution and complementarity for optimal pricing.

1. Introduction

The choice environment facing consumers has become increasingly complex. Consumers today often have the option of creating bundles of attributes, products, and services to construct their choices rather than picking from a predefined set of alternatives. Such "menu-based" choice environments, once considered rather unique, have become popular in a number of industries and product categories. The classic example of such an environment is the construction of a restaurant order for which one might have the choice of a starter, a main course, and a dessert from an actual menu. Ultimately, the utility-maximizing choice from such a menu is some combination of the individual items available in the choice set. Any approach to modeling the revealed choice of the consumer, then, would need to accommodate the many combinations of substitutes, complements, and independent items implicit in the menus.

Such menus (apart from the restaurant setting) are now found in purchase contexts across a variety of industries. Examples include the personal computing industry in which products are almost always built based on consumers' choice of individual components (memory, storage, processor, screen, etc.), most electronic devices (tablets, phones, etc.), and more recently, products and service industries as diverse as financial products (vanguard.com), sneakers (Nike, Adidas, Converse), furniture (Herman Miller), household products (levolor.com, e-skylight.com), clothing (make-yourownjeans.com, jeanuine.com), pizza, and even muesli (mymuesli.com). Such menu-based choice settings have also been used (under the name menu-based choice experiments) to elicit stated preferences of consumers in experimental settings (Liechty et al. 2001, Orme 2010). The prototypical menu-based choice experiment (henceforth MBCE) presents each respondent with multiple menus from which to choose, varying the prices or the availability of individual items in the menus (Orme 2010). The resulting data are used to calibrate choice models designed to rationalize the choice of the various combinations of individual items[1] from the menus.

As one would expect, the fundamental difference between traditional choice settings and menu-based choices is that the items constituting the choice might interact with each other in ways that make them substitutes or complements. The interest in modeling choices from menus and the interactions between individual items constituting these is not new (see, e.g., Farquhar and Rao 1976, Chung and Rao 2003, and Gentzkow 2007).

However, models and applications in this domain typically included a limited number of possible item combinations and, consequently, were able to compute the choice probabilities that spanned all possible combinations stemming from the menus under investigation directly. As the set of items available becomes large, the power set spanned by the menu (i.e., the complete set of possible combinations) increases exponentially, rendering the full enumeration approach (to computing choice probabilities) difficult if not impossible. In other words, the menu-based choice problem is often intractable and more so when taking heterogeneity into account. To see this, consider a menu with 20 binary items, not atypical in modern business applications. The consumer then can construct more than a million combinations. As one moves to even slightly larger menus, the power set explodes (for 30 items, we have more than a billion combinations).

A different route to the menu choice problem is taken in Liechty et al. (2001) and Orme (2010). Instead of estimating probabilities for choices from a menu directly, Liechty et al. (2001) and Orme (2010) estimate (conditional) probabilities of choosing individual items from the menu. The probability of a menu choice is, thus, defined implicitly, breaking the curse of dimensionality from the exploding number of possible item combinations in large menus. Liechty et al. (2001) estimate conditional distributions derived from a multivariate probit model with correlated error terms. Error correlations are intended to capture substitutive and complementary relations between items in menus. Orme (2010) estimates independent binomial logit models for each item and includes selected cross-price effects in these models to capture demand interdependencies between items.

Although both the approaches can be applied to large menus, we show that correlated errors as well as cross-price effects included in individual-level choice models only provide limited accounts of substitution and complementarity because they fail to generate important demand implications by these economic concepts.

Our paper takes a different approach to resolving the intractability of menu choice problems. Following ideas laid out in Besag (1972, 1974) and first introduced to marketing by Russell and Petersen (2000), we propose a choice framework that, in principle, allows for arbitrary demand interdependencies across items in the choice environment as well as heterogeneity in the structural parameters of interest.

Although similar ideas have been proposed before (see, e.g., Russell and Petersen 2000 and Kamakura and Kwak 2012), the implementation of such a framework to real data with unobserved heterogeneity has been a challenge. This paper outlines a Bayesian inference approach to operationalizing the proposed framework that *exactly* corresponds to the likelihood of our menu-choice model. The proposed estimation framework does not rely on any ad hoc approximations or heuristics to approximate the likelihood despite the fact that the direct evaluation of this likelihood is intractable in large menus. Our approach stands in contrast to earlier attempts to tackle this problem using pseudo-likelihoods or by resorting to approximate Monte Carlo approaches (e.g., Kamakura and Kwak 2012).

In particular, we develop Bayesian inference for a hierarchical version of the model that includes possibly informative priors supporting strong utility dependencies, for example, essentially perfect substitution among a subset of items, but allows for individual-level departures. These individual consumer-level departures are examples of extreme heterogeneity that arises naturally as a function of what different consumers perceive to be (only) vertically or horizontally differentiated items in a menu.[2]

Estimation in this framework remains complicated given that the likelihood continues to be intractable. We then propose a new adaptation of the exchange algorithm proposed by Murray et al. (2006) generalizing it to a setting with heterogeneous decision makers. To the best of our knowledge, this is the first adaptation of the exchange algorithm for Bayesian inference in the setting of a hierarchical model that accounts for unobserved heterogeneity.

Our approach allows us to construct good proposal densities based on approximate data augmentation ideas that nevertheless maintain the detailed balance conditions. The method facilitates Metropolis–Hastings sampling in situations in which large step sizes are required by construction of the model. Because the exchange algorithm requires an exact sampler for augmentation purposes, we construct a perfect sampler using the coupling from the past paradigm proposed by Propp and Wilson (1996). The combination of these various subalgorithms circumvents the evaluation of the normalizing constant in the likelihood that is computationally intractable in large choice sets. Importantly, the exact sampler we employ in the exchange-step of our algorithm guarantees that our adaptation of the exchange algorithm converges to the *exact* posterior implied by our model and prior.[3]

The novelty of our contribution is, thus, that we make the model developed in Besag (1972, 1974) and

first introduced to marketing by Russell and Petersen (2000) operational for marketing applications in which heterogeneity is a first-order concern and beyond the limits of small menus. To the best of our knowledge, the account of heterogeneity in utility interactions we propose is entirely new and turns out to be important for applications in which some consumers view (a subset of items) as essentially perfect substitutes, whereas other consumers appreciate horizontal differences between the same items. Furthermore, we show, based on first principles, that the core model underlying this framework has important advantages over the feasible extant models by Liechty et al. (2001) and Orme (2010) in the way it measures substitution and complementarity between items in a menu. Specifically, we show that correlated errors as well as cross-price effects only provide a limited representation of complementarity and substitution between items in a menu with counterintuitive implications for optimal actions.

We apply our menu-choice model to data from an MBCE designed to study demand for game consoles and accessories by GfK. We find empirical evidence for the relevance of utility dependencies because of substitution and complementarity in our data, demonstrate superior predictive performance relative to extant models, and illustrate implications for optimal pricing of items and bundles in a menu.

In Section 2, we review the menu-choice model and contrast it to the multivariate probit model (Liechty et al. 2001) and to including cross-price effects in independent binomial logit models (Orme 2010) in Section 3. In Section 4, we discuss Bayesian inference and develop our hierarchical framework. A discussion of the data and estimation results is in Section 5. Section 5.5 discusses managerial implications. We conclude with a summary and an outlook.

2. A Model for Menu-Based Choice

Consider a menu containing $k = 1, \ldots, K$ items and denote a (multivariate) choice that consumer i makes as

$$\mathbf{Y}_i = \{Y_{i,1}, \ldots, Y_{i,k}, \ldots, Y_{i,K}\},$$

where Y_{ik} takes the value 1 (0) if consumer i's choice includes (does not include) item k. With K items in a menu, there are $j = 1, \ldots, J$ possible choices, where $J = 2^K$. The complete set of these 2^K possible choices is written as $\mathcal{Y}$. This includes choosing nothing, that is, the outside good from the menu. We express the utility from some choice $\mathbf{Y}_i$ as

$$U(\mathbf{Y}_i) = U(\mathbf{Y}_i; \mathbf{X}, \Psi_i) + \varepsilon(\mathbf{Y}_i),$$

where $U(\mathbf{Y}_i; \mathbf{X}, \Psi_i)$ first is an essentially arbitrary characterization of the indirect utility associated with the

choice $\mathbf{Y}_i = \{Y_{i,1}, \ldots, Y_{i,K}\}$ with Ψ_i denoting preference parameters (to be estimated) and $\mathbf{X}$ covariates associated with the choice $\mathbf{Y}_i$. Finally, $\varepsilon(\mathbf{Y}_i)$ is an idiosyncratic error component. Assuming the idiosyncratic error ε to be independent and identically distributed extreme value type I, the probability of choice $\mathbf{Y}_i$ is given by the multinomial logit model (MNL) over the set of possible choices $\mathcal{Y}$ in Equation (1).

$$\Pr(\mathbf{Y}_i) = \frac{\exp(U(\mathbf{Y}_i; \mathbf{X}, \Psi_i))}{\sum_{Y' \in \mathcal{Y}} \exp(U(\mathbf{Y}'_i; \mathbf{X}, \Psi_i))} = \frac{\exp(U(\mathbf{Y}_i; \mathbf{X}, \Psi_i))}{\mathcal{Z}(\mathbf{X}, \Psi_i)}.$$
(1)

Here, $\mathcal{Z}(\mathbf{X}, \Psi_i)$ is compact notation for the denominator $\sum_{Y' \in \mathcal{Y}} \exp(U(\mathbf{Y}'_i; \mathbf{X}, \Psi_i))$ of this MNL model.[4]

This MNL model is consistent with utility maximization. The usual description of the process that generates the MNL probabilities involves the consumer evaluating the utilities of each alternative (in this case, each of the 2^K possible choices) and then picking the one that offers the maximum utility. With the number of possible choices exploding in large menus, this may be an improbable assumption about the generative process.

Next, we show that the MNL probabilities emerge as the equilibrium distribution (over the space of possible choices $\mathcal{Y}$) of a simple, iterative stochastic decision process. Although we do not claim that this is the precise decision process that generates the probabilities defined in Equation (1), it does point to the existence of alternative generative models and, thus, decision processes that result in logit-based choice models. Specifically, we show that the choice probabilities in Equation (1) can result from a decision process that does not require the enumeration of all possible 2^K choices in $\mathcal{Y}$.

2.1. A Generative Logit Model of Menu-Based Choice

Consider, as before, a consumer faced with the decision to choose from a menu comprised of K items. As such, there are $j = 1, \ldots, J$ possible choice alternatives, where $J = 2^K$. Assume also, as before, that the deterministic utility of a bundle is described as earlier by $U(\mathbf{Y}_i; \mathbf{X}, \Psi_i)$.

The consumer starts the decision process with some arbitrary choice at time epoch $s = 0$,[5] say $\mathbf{Y}_i^{(0)} = \{Y_{i,1}^{(0)}, \ldots, Y_{i,K}^{(0)}\}$. The consumer then focuses attention on a given item k to evaluate. We assume that, at any time s, there is a positive probability $\lambda_{ik} > 0$ that item k is the focus of attention. These attention probabilities can take a fairly arbitrary form; for example, they may depend on preference parameters (Ψ_i), item characteristics ($\mathbf{X}$), or be constant, say with $\lambda_{ik} = \lambda = \frac{1}{K}$. The only requirements are that the probability of considering any item at any time be strictly greater than zero,

that they sum to one, and that they not depend on the current instantiation of the choice (i.e., not depend on $Y_i^{(s)}$). Once the consumer has focused attention on a given item, the consumer decides whether to include or exclude item k in the choice. To do so, we assume that the consumer can describe the consumer's utility for any choice both with and without the focal item as in Equation (2), where the notation $\mathbf{Y}_{i,-k}$ is short for $\mathbf{Y}_i$ excluding the kth item.

$$
\begin{aligned}
U\left(Y_{i,k}^{(s)} = 0, \mathbf{Y}_{i,-k}^{(s-1)}; \mathbf{X}, \Psi_i\right) + \epsilon_{0,i,k}^{(s)} \\
U\left(Y_{i,k}^{(s)} = 1, \mathbf{Y}_{i,-k}^{(s-1)}; \mathbf{X}, \Psi_i\right) + \epsilon_{1,i,k}^{(s)}.
\end{aligned}
\tag{2}
$$

The decision rule then to include (or keep) item k is based on utility maximization. Put simply, the consumer is evaluating whether the addition or removal of an item is utility-enhancing given the current state of the choice vector. Item k is included in the choice if including it increases utility as compared with excluding it, that is, the inequality in Equation (3) holds.

$$
\begin{aligned}
U\left(Y_{i,k}^{(s)} = 1, \mathbf{Y}_{i,-k}^{(s-1)}; \mathbf{X}, \Psi_i\right) + \epsilon_{1,i,k}^{(s)} \\
> U\left(Y_{i,k}^{(s)} = 0, \mathbf{Y}_{i,-k}^{(s-1)}; \mathbf{X}, \Psi_i\right) + \epsilon_{0,i,k}^{(s)}.
\end{aligned}
\tag{3}
$$

The conditional probability reflecting the decision regarding item k at epoch s is then in the form of the binomial logit model in Equation (4).

$$
\begin{aligned}
&\Pr\left(Y_{i,k}^{(s)} = 1 | \mathbf{Y}_{i,-k}^{(s-1)}, \mathbf{X}, \Psi_i\right) \\
&= \frac{\exp\left(U\left(Y_{i,k}^{(s)} = 1, \mathbf{Y}_{i,-k}^{(s-1)}; \mathbf{X}, \Psi_i\right)\right)}{\begin{aligned}&\exp\left(U\left(Y_{i,k}^{(s)} = 0, \mathbf{Y}_{i,-k}^{(s-1)}; \mathbf{X}, \Psi_i\right)\right) \\ &+ \exp\left(U\left(Y_{i,k}^{(s)} = 1, \mathbf{Y}_{i,-k}^{(s-1)}; \mathbf{X}, \Psi_i\right)\right)\end{aligned}}.
\end{aligned}
\tag{4}
$$

If this process continues, the consumer exhibits the choice vector $\mathbf{Y}_i$ with probability exactly equal to the MNL model from Equation (1). A formal proof of this claim is provided in Online Appendix B. The result obtains from recognizing that the process described is a Markov chain with transition probabilities defined by the conditional choice probabilities and the attention probabilities λ_{ik}. Further, because all states communicate, the sequence $[\mathbf{Y}_i^{(s)}]_{s=0}^{S}$ is irreducible, and aperiodicity can be verified straightforwardly. Because the state space is finite, this Markov chain is consequently ergodic and has a unique stationary distribution. We show in Online Appendix B that the stationary distribution under which detailed balance holds is described by Equation (1).

Under this framework, if $\mathbf{Y}^{(s)}$ is drawn from the equilibrium distribution, then any element-wise operations, as long as they are done in a utility-theoretic manner, result in a draw $\mathbf{Y}^{(s+1)}$ from the same probability distribution. One interpretation of utility maximization, thus, is that a consumer has cycled through the conditional distributions in Equation (4) until convergence to the equilibrium in Equation (1), presumably starting from $\mathbf{Y}_i^{(0)} = \{0, \ldots, 0\}$. We informally note that—in larger menus—convergence to a $\mathbf{Y}$ distributed according to Equation (1) following this process typically occurs in a small number of steps relative to the $J = 2^K$ possible item combinations. This is because, following this generative process, the consumer likely arrives at a $\mathbf{Y}$ distributed according to Equation (1) before considering, that is, explicitly evaluating, the potentially huge number of unattractive, low-utility item combinations.

We repeat here that we are not claiming or suggesting that this is the exact process used by consumers. Similarly, we cannot claim that enumerating all combinations and choosing the maximum is the process by which the logit probabilities derive. Each of these processes is a way to rationalize the choice model. The purpose of our exercise is simply to demonstrate that consumers need not solve a very complicated maximization process to generate choices that can be rationalized by the logit model over the set $\mathcal{Y}$ of all possible choices that explodes in large menus.

We would like to point out here that these processes (maximization, pairwise maximization) are not the only approaches to realizing the logit choice probabilities. An alternative derivation of the logit model is due to Besag (1972), who originally proposed this model for the analysis of multivariate binary spatial data and termed this the autologistic model.

Besag's (1972) derivation is based on a causal argument about the nature of interactions between sites on a spatial array. The argument states that the conditional probability of an event at a focal site in the array only depends on observable states of directly connected neighboring sites, that is, on whether an event has occurred at these sites or not, and possibly on nonspatial characteristics of the focal site. In our context, the array is the menu from which a respondent chooses, events on the array are choices of individual items from the menu, and the connections between items are due to complementarity or substitution. In a sense, Besag's (1972) arguments directly lead to the detailed specification of $U(\mathbf{Y}_i; \mathbf{X}, \Psi_i)$ that we discuss next.

2.2. Parametric Assumptions and Implied Model Properties

The menu-choice model (MCM) described is general. In order to take the model to data, we need to spell out the exact specifications for the utility $U(\mathbf{Y}_i; \mathbf{X}, \Psi)$. In what follows, we state and briefly discuss our assumptions and specification choices.

Assumption 1. *We assume that the utility the consumer obtains from a bundle of items constituting the consumer's choice depends only on the items contained in that bundle. This assumption is common to all choice models rooted in utility theory, in which the direct utility of an option cannot depend on characteristics of other options. It follows that only the characteristics (including prices) of those items included in the bundle $\mathbf{Y}_i$ matter in $U(\mathbf{Y}_i; \mathbf{X}, \Psi)$.*

Assumption 2. *We assume that utility is additively separable in the main and interaction effects of the items. As we show later in more detail, this assumption leads to particularly simple expressions for conditional item choice probabilities, reflective of Besag's (1972) causal arguments. That is, conditional item choice probabilities only depend on the specifics of an item and—through utility interactions—on what other items have been chosen.*

Assumption 3. *We assume that interactions between items in a bundle can be constructed as the sum of pairwise item interactions. This assumption reduces the number of interaction parameters in the utility function to the number of unique pairs of items in the complete menu, that is, $(K^2 - K)/2$. Assumption 3 is mostly for pragmatic reasons. Even after imposing the structurally motivated restrictions in Assumptions 1 and 2, estimating all possible higher order utility interactions among items would leave us with $2^K - K - 1$ interaction parameters to estimate.[6] In all but the smallest menus, an attempt to calibrate all such model parameters from observed choices is hopeless because the number of observations is likely smaller than the number of parameters. The specification we adopt reduces the number of parameters to estimate substantially and is still extremely flexible in that every item potentially interacts with every other item in the menu. However, it rules out higher order interactions. An example for a higher order interaction are three individual items that only become attractive when chosen together.*

Together, these assumptions translate into utilities of the form given in Equation (5) that correspond to the multivariate logistic model first introduced to marketing by Russell and Petersen (2000).

$$U(\mathbf{Y}_i; \mathbf{X}, \Psi_i) = \sum_{k=1}^{K} \beta_{i,k} Y_{i,k} + \sum_{k=1}^{K} \sum_{k'=k+1}^{K} \theta_{k,k'} Y_{i,k} Y_{i,k'} \quad (5)$$

$$= \mathbf{Y}_i' \boldsymbol{\beta}_i + \mathbf{Y}_i' \Theta_i \mathbf{Y}_i$$

$$\Theta = \begin{pmatrix} 0 & \theta_{1,2} & \theta_{1,3} & \theta_{1,4} & \dots & \theta_{1,K} \\ 0 & 0 & \theta_{2,3} & \theta_{2,4} & \dots & \theta_{2,K} \\ 0 & 0 & 0 & \theta_{3,4} & \dots & \theta_{3,K} \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \quad (6)$$

Here, preference parameters are $\Psi_i = \{\boldsymbol{\beta}_i, \Theta_i\}$, where β_k is a parameter reflecting the attractiveness of or utility

from item k, and the entries in the upper triangular matrix Θ measure the strength of complementary and substitutive relations between the items.

Many MBCEs include utility shifters, such as, for example, prices of individual items that vary across menus as part of an experimental design. We follow the standard practice of including price as a "linear attribute" in the (indirect) utility of individual items:

$$\beta_k = \beta_{k,0} + \beta_{price} p_k, \quad (7)$$

Equation (7) implies a quasilinear utility specification for choices $j = 1, \dots, J$ from a menu (see Equation (5)) from which income drops out. Therefore, the utility interactions $\theta_{k,k'}$ should be interpreted as interactions in direct utility. For example, a negative element $\theta_{k,k'}$ is negative because alternative k makes alternative k' redundant independent of the income consumed by purchasing k.

The specification we have adopted implies some useful properties. These include the decomposition of the MNL into a series of tractable and closed-form item-level conditional choice probabilities and intuitive demand dependencies between (sets of) menu items. We discuss these starting with the simplest case of a two-item menu and deferring general results to Online Appendix A.

In a two-item menu with items A and B and item-specific utilities a and b as well as the utility interaction parameter θ_{AB}, the probabilities of choosing only A, only B, $\{A, B\}$, and the outside good are proportional to $\exp(a)$, $\exp(b)$, $\exp(a + b + \theta_{AB})$, and $\exp(0)$, respectively, with the normalizing constant equal to the sum of these four terms (from Equations (1) and (5)). According to Equation (4) and following basic principles of probability, the conditional probability of choosing A given that B is chosen is then

$$\begin{aligned} \Pr(Y_A = 1 | Y_B = 1) &= \frac{\Pr(Y_A = 1, Y_B = 1)}{\Pr(Y_B = 1)} \\ &= \frac{\exp(a + b + \theta_{AB})}{\exp(b) + \exp(a + b + \theta_{AB})} \\ &= \frac{\exp(a + \theta_{AB})}{1 + \exp(a + \theta_{AB})}, \end{aligned} \quad (8)$$

where $\Pr(Y_B = 1) = \Pr(Y_B = 1, Y_A = 0) + \Pr(Y_B = 1, Y_A = 1)$. Thus, under this specification, the conditional probability of including item A in the bundle is in the form of a binomial logit model that only depends on the utility of A and the interaction parameter but not on the item-specific utility of item B reflecting Besag's (1972) original causal argumentation. By the same basic probability argument, we have that

$$\Pr(Y_A = 1 | Y_B = 0) = \frac{\exp(a)}{1 + \exp(a)}. \quad (9)$$

Again this probability is independent of the item-specific utility of B.

If item B neither enhances nor subtracts from the utility of A and vice versa, that is, if there is no utility connection between A and B, then $\theta_{AB} = 0$ and $\Pr(Y_A = 1|Y_B = 1) \equiv \Pr(Y_A = 1|Y_B = 0)$. If there is a connection between A and B such that B substitutes for A (and vice versa), then $\theta_{AB} < 0$, and consequently, the probability of choosing A given $Y_B = 1$ is smaller than the probability of choosing A as a stand-alone item. Finally, if item B complements A (and vice versa), then $\theta_{AB} > 0$ and $\Pr(Y_A = 1|Y_B = 1) > \Pr(Y_A = 1|Y_B = 0)$.

Finally, we show that the cross-derivative of the *marginal* probability of choosing item A with respect to the attractiveness of item B depends on the interaction term θ_{AB} as well and that standard multinomial logit choice among *individual items*, that is, the standard MNL model, arises as a special case of the MCM. The marginal probability of choosing A from this menu is the sum of the probabilities of choosing only A and of choosing A and B together (Equation (10)).

$$\Pr(Y_A = 1) = \frac{\exp(a) + \exp(a + b + \theta_{AB})}{1 + \exp(a) + \exp(b) + \exp(a + b + \theta_{AB})}. \tag{10}$$

Multinomial logit choice among individual items in a choice set results from this equation when utility interactions (θ_{AB} in this example) equal negative infinity. That is, multinomial logit choice among individual items in a choice set corresponds to the situation in which item combinations are completely dominated by choices of individual items because of perfect substitution between individual items.[7]

Taking the derivative of Equation (10) with respect to the item B–specific utility b, we get

$$\frac{\partial \Pr(Y_A = 1)}{\partial b} = \frac{\left(1 + e^a + e^b + e^{a+b+\theta_{AB}}\right)e^{a+b+\theta_{AB}} - \left(e^b + e^{a+b+\theta_{AB}}\right)\left(e^a + e^{a+b+\theta_{AB}}\right)}{\left(1 + e^a + e^b + e^{a+b+\theta_{AB}}\right)^2},$$

which simplifies to Equation (11).

$$\frac{\partial \Pr(Y_A = 1)}{\partial b} = \frac{e^{a+b+\theta_{AB}} - e^{a+b}}{\left(1 + e^a + e^b + e^{a+b+\theta_{AB}}\right)^2}. \tag{11}$$

Then it is straightforward to see that

$$\text{sign}\left(\frac{\partial \Pr(Y_A = 1)}{\partial b}\right) = \text{sign}(\theta_{AB}).$$

In other words, the cross-derivative of the marginal probability of choosing A with respect to the attractiveness of B is positive (negative) when $\theta_{AB} > 0$ ($\theta_{AB} < 0$), that is, when B complements (substitutes) A. When $\theta_{AB} > 0$, increasing the attractiveness of B increases the

marginal probability of choosing A. Obviously, if $\theta_{AB} < 0$, substitution effects obtain. We characterize conditional and marginal probabilities and cross-derivatives in general menus in Online Appendix A.

In the next section, we contrast the existing feasible models to the menu choice model just characterized with a special emphasis on how these models account for substitution and complementarity between items in a menu.

3. Extant Models
3.1. The Multivariate Probit Model
Liechty et al. (2001) proposed the multivariate probit model (MvP) as basic model for data from MBCEs with the understanding that error correlations capture complementary and substitutive relationships between items in the menu. However, although error correlations in the MvP can certainly *reflect* dependencies from complementarity or substitution, the MvP fails at *generating* important implications from the economics of substitution and complementarity. Most notably, all cross-derivatives in the MvP and, thus, all cross-price elasticities are zero by definition of the model if the only connection between item-specific utilities is through correlated errors.

This property of the MvP is best illustrated using an example. Consider a menu consisting of the two alternatives A and B with attractiveness or utility a and b, and define the corresponding random utilities as z_A and z_B. Then the marginal probability of choosing A from this two-item menu is defined as $\Pr(z_A > 0)$. Starting from the joint distribution $p(z_A, z_B)$, this probability corresponds to the following double integral:

$$\Pr(Y_A = 1)$$
$$= \int_{-\infty}^{\infty} \int_0^{\infty} N\left(\begin{bmatrix} z_A \\ z_B \end{bmatrix} \middle| \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} 1 & \rho_{AB} \\ \rho_{AB} & 1 \end{bmatrix}\right) dz_A dz_B$$
$$= \int_0^{\infty} N(z_A|a, 1) dz_A, \tag{12}$$

where ρ_{AB} is the correlation between random utilities z_A and z_B. As is obvious from the last line of Equation (12), $\partial \Pr(Y_A = 1)/\partial b = 0$; that is, the marginal share of A is independent of the attractiveness of B for all latent utility correlations ρ_{AB}. This implication of the MvP is inconsistent with the economics of substitution and complementarity that imply decreasing (increasing) demand for the substitute (complement) A as the attractiveness of B increases. It follows that, in applications in which a prior understanding of the items in a menu suggest that cross-derivatives are likely to be nonzero, for example, that increasing the price of item B shifts demand from B to A, an account that is based on correlated errors only is inadequate (cf. Equation (11)).[8]

A related question is how to empirically distinguish between dependencies that arise from utility correlations as in the MvP and utility interactions as in the MCM. We noted earlier that repeated choices from a fixed menu are sufficient to just identify MCM parameters after imposing our Assumptions 1 and 2 (see Section 2.2). In a menu comprised of only the two items A and B, the data yield four probabilities: $\Pr(A)$, $\Pr(B)$, $\Pr(A,B)$, and the probability of choosing the outside good denoted as $\Pr(\emptyset)$. These probabilities identify three linearly independent parameters. Thus, before invoking special cases, such as, for example, perfect substitution between two individually attractive items, both the MCM and the MvP are just identified and, thus, empirically indistinguishable in this example.

However, exogenous variation in the baseline probability of A or B distinguishes between the MCM and the MvP. Consider, for example, choices from three different menus consisting of $\{A\}$, $\{B\}$, and $\{A,B\}$, respectively, and note that the probability of choosing, for example, B from the menu consisting of only $\{A\}$ is exogenously equal to zero. According to the MvP, the marginal probability of choosing, for example, A is independent of the menu context, that is, $\Pr(Y_A = 1|\{A\}) = \Pr(Y_A = 1|\{A,B\})$; see Equation (12). In contrast, the MCM predicts $\Pr(Y_A = 1|\{A\}) \neq \Pr(Y_A = 1|\{A,B\})$ unless there is no utility interaction between items A and B, that is, $\theta_{AB} = 0$ (see Equations (10) and (11)). The same argument holds if there are utility shifters that vary the attractiveness of individual items in a menu exogenously.

Another related question is if correlations between latent utilities as in the MvP and utility interactions as in the MCM can be jointly identified. We defer a thorough answer to this question to future research. Gentzkow (2007), who estimates both utility interactions and correlations for the two-alternative case of the MCM described in Section 2, offers some discussion. In applications to MBCEs, substitution and complementarity together with persistent heterogeneity are the first-order concerns. We leave the discussion of direct utility interactions versus utility correlations at the level of model comparisons.

3.2. Independent Logits with Cross-Price Effects

The current industry standard to handle demand dependencies among items in MBCEs is to include selected cross-price effects (CPEs) into the items' utility functions (Orme 2010). To continue the example of a menu comprised of items A and B, the random utilities of items A and B are specified as

$$u_A = \beta_{A,0} + \beta_{price}p_A + \beta_{cp}^{(B \to A)}p_B + \epsilon_A, \ \epsilon_A \sim Logistic$$

$$u_B = \beta_{B,0} + \beta_{price}p_B + \beta_{cp}^{(A \to B)}p_A + \epsilon_B, \ \epsilon_B \sim Logistic,$$

$$(13)$$

resulting in independent binomial logit models for items in a menu (henceforth IndepCPE). Here the "cross-price" (cp) parameter $\beta_{cp}^{(B \to A)}$ ($\beta_{cp}^{(A \to B)}$) captures the change in A's (B's) indirect utility as a function of changes in the respective other brand's price.

The intuition is that positive cross-price effects capture substitution, and negative cross-price effects complementarity. This intuition has been honed in aggregate demand models (e.g., Carpenter et al. 1988, Blattberg and Wisniewski 1989, Hoch et al. 1995, and Wedel and Zhang 2004) in which cross-price effects capture the change in the aggregate sales or market share of a focal brand as a function of other brands' price changes, and positive (negative) cross-price coefficients have been interpreted as aggregate measures of substitution (complementarity) between brands.

However, on the right-hand side in Equation (13), we have utilities and not aggregate sales or shares. Although the distribution of utilities implies choice probabilities and choice probabilities aggregate to market shares, the relationship between coefficients in the individual utility function and coefficients in an aggregate model is not straightforward.

Consider the case of perfect substitution between about equally attractive items A and B. As described following Equation (10) in Section 2.2, perfect substitution between items A and B corresponds to multinomial choice from the set $\{A, B, \emptyset\}$. At fixed prices that yield about equal indirect utilities for items A and B in this example, the resulting choice probabilities of A and B are about equal, that is, $\Pr(A) = \Pr(B)$. However, because of perfect substitution, the probability of choosing A and B together is zero. Fitting data from this process using the IndepCPE defined in Equation (13) faces the problem that the fitted utility functions will *necessarily* predict a positive probability for choosing both A and B together that is equal to the product of $\Pr(A)$ times $\Pr(B)$. Thus, whenever the shares of A and B are large, relative to the outside good, the best approximation based on Equation (13) predicts a substantial share of choosing A and B together. For the same structural reason, IndepCPE tends to underestimate the probability of choosing complements together. The bias is more pronounced for individually less attractive alternatives A and B when the product of $\Pr(A)$ times $\Pr(B)$ is small. Thus, IndepCPE fails at generating substitution and complementarity between items.

In addition, cross-price effects may even fail to correctly reflect the sign of utility interactions, depending on the amount of price variation in the data (see Online Appendix E for numerical illustrations). Last but not least, the implication from IndepCPE that item utilities depend on the prices of other items in the menu

contradicts basic utility theory. As a practical consequence, optimization over prices given parameters is bound to yield solutions that suggest maximizing demand for a high-margin item by setting the price for other items to extreme values for common functional forms of the cross-price effects.

We conclude that the MCM has important theoretical advantages over the MvP and independent binomial logit models that include cross-price effects (IndepCPE) as a basic model for MBCEs. Next, we develop Bayesian inference for this model, including a parsimonious prior for the utility interaction parameters in the MCM.

4. Bayesian Inference
The MCM described in Section 2.2 implies the following individual i $(i = 1, \ldots, N)$-specific indirect utility function for choice consisting of some combination of items from menu t $(t = 1, \ldots, T)$ with a total of K items:

$$U(\mathbf{Y}_{i,t}; \mathbf{X}_{i,t}, \Psi_i) = \sum_{k=1}^{K} Y_{i,t,k}(\beta_{i,k} + \beta_{i,price}p_{t,k}) + \mathbf{Y}'_{i,t}\Theta_i\mathbf{Y}_{i,t}. \quad (14)$$

Inference for the parameters in Equation (14) is challenging, in particular in the context of MBCEs. First, the normalizing constant of the likelihood in Equation (1) has $J = 2^K$ terms and becomes prohibitively expensive to compute in larger menus. We propose a solution to this problem enabling likelihood-based Bayesian inference based on the recently developed exchange algorithm (Möller et al. 2006, Murray et al. 2006). Second, in MBCEs, each of N respondents typically only make choices from a small set of different menus; that is, T is small and $N >> T$. Therefore, it is natural to attempt inference using a hierarchical model that pools information across respondents. Next, we briefly describe the exchange algorithm and then introduce our hierarchical prior.

4.1. The Exchange Algorithm
Define $\mathbf{Y}_i = (\mathbf{Y}'_{i,1}, \ldots, \mathbf{Y}'_{i,T})$ as the matrix of choices by individual i from T different menus, in which $Y_{i,t,k} = 1$ if individual i chooses item k from menu t and else zero.

Denote by $p(\Psi_i|\mathbf{Y}_i)$ the *normalized posterior*, $\Pr(\mathbf{Y}_{i,t}|\Psi_i)$ the normalized *likelihood* defined by Equations (1) and (14) with normalizing constant $\mathscr{L}_t(\Psi_i)$ at the tth observation,[9] $\ell^*(\mathbf{Y}_{i,t}|\Psi_i)$ the *non*-normalized likelihood (the numerator in Equation (1)), $p(\Psi_i)$ as the prior

and $q(\Psi_i)$ as the proposal density.[10] Then the exchange algorithm uses the following acceptance probability:

$$\alpha_{exchange}(\Psi_i \rightarrow \Psi_i^c, \mathbf{Y}_{i,t}^c)$$

$$= \min\left(1, \frac{p(\Psi_i^c)q(\Psi_i)}{p(\Psi_i)q(\Psi_i^c)} \prod_{t=1}^{T} \frac{\Pr(\mathbf{Y}_{i,t}|\Psi_i^c)}{\Pr(\mathbf{Y}_{i,t}|\Psi_i)} \cdot \frac{\Pr(\mathbf{Y}_{i,t}^c|\Psi_i)}{\Pr(\mathbf{Y}_{i,t}^c|\Psi_i^c)}\right)$$

$$= \min\left(1, \frac{p(\Psi_i^c)q(\Psi_i)}{p(\Psi_i)q(\Psi_i^c)} \prod_{t=1}^{T} \frac{\ell^*(\mathbf{Y}_{i,t}|\Psi_i^c)}{\mathscr{L}_t(\Psi_i^c)}\right.$$

$$\left. \cdot \frac{\mathscr{L}_t(\Psi_i)}{\ell^*(\mathbf{Y}_{i,t}|\Psi_i)} \frac{\ell^*(\mathbf{Y}_{i,t}^c|\Psi_i)}{\mathscr{L}_t(\Psi_i)} \frac{\mathscr{L}_t(\Psi_i^c)}{\ell^*(\mathbf{Y}_{i,t}^c|\Psi_i^c)}\right)$$

$$= \min\left(1, \frac{p(\Psi_i^c)q(\Psi_i)}{p(\Psi_i)q(\Psi_i^c)} \prod_{t=1}^{T} \frac{\ell^*(\mathbf{Y}_{i,t}|\Psi_i^c)}{\ell^*(\mathbf{Y}_{i,t}|\Psi_i)} \cdot \frac{\ell^*(\mathbf{Y}_{i,t}^c|\Psi_i)}{\ell^*(\mathbf{Y}_{i,t}^c|\Psi_i^c)}\right)$$

$$= \min\left(1, \frac{p(\Psi_i^c)q(\Psi_i)}{p(\Psi_i)q(\Psi_i^c)} \prod_{t=1}^{T} \frac{\ell^*(\mathbf{Y}_{i,t}|\Psi_i^c)}{\ell^*(\mathbf{Y}_{i,t}|\Psi_i^c)} \cdot \frac{\ell^*(\mathbf{Y}_{i,t}^c|\Psi_i)}{\ell^*(\mathbf{Y}_{i,t}|\Psi_i)}\right), \quad (15)$$

where $\mathbf{Y}_{i,t}^c \sim \Pr(\mathbf{Y}_{i,t}^c|\Psi_i^c)$ is obtained using a *perfect* sampler (Propp and Wilson 1996) and *without* evaluating the normalized likelihood; see Online Appendix C for details.[11]

Expanding the state space from Ψ_i to include $\mathbf{Y}_i^c$, the exchange algorithm circumvents the evaluation of $\mathscr{L}_1(\Psi_i), \ldots, \mathscr{L}_T(\Psi_i)$ and $\mathscr{L}_1(\Psi_i^c), \ldots, \mathscr{L}_T(\Psi_i^c)$ that cancel from Equation (15). The algorithm is called an *exchange* algorithm because it combines the move from $\mathbf{Y}_i$ to $\mathbf{Y}_i^c$ given the *old* parameter vector Ψ_i with the "move back" (to the actual data) from $\mathbf{Y}_i^c$ to $\mathbf{Y}_i$ given the *new* parameter vector Ψ_i^c and, thus, toward the proposed parameter (see the last line of Equation (15)). It, thus, exchanges $\mathbf{Y}_i$ with $\mathbf{Y}_i^c$ given Ψ_i and $\mathbf{Y}_i^c$ with $\mathbf{Y}_i$ given Ψ_i^c when moving from Ψ_i to Ψ_i^c. We obtain the "traditional" marginal MH-acceptance ratio by integrating out $\mathbf{Y}_i^c$:

$$\alpha_{marginal}(\Psi_i \rightarrow \Psi_i^c)$$

$$= \min\left(1, \frac{p(\Psi_i^c)q(\Psi_i)}{p(\Psi_i)q(\Psi_i^c)} \prod_{t=1}^{T} \int \frac{\ell^*(\mathbf{Y}_{i,t}|\Psi_i^c)}{\ell^*(\mathbf{Y}_{i,t}|\Psi_i)} \frac{\ell^*(\mathbf{Y}_{i,t}^c|\Psi_i)}{\ell^*(\mathbf{Y}_{i,t}^c|\Psi_i^c)}\right.$$

$$\left. \cdot \frac{\ell^*(\mathbf{Y}_{i,t}^c|\Psi_i^c)}{\mathscr{L}_t(\Psi_i^c)} d\mu(\mathbf{Y}_{i,t}^c)\right)$$

$$= \min\left(1, \frac{p(\Psi_i^c)q(\Psi_i)}{p(\Psi_i)q(\Psi_i^c)} \prod_{t=1}^{T} \frac{\ell^*(\mathbf{Y}_{i,t}|\Psi_i^c)}{\ell^*(\mathbf{Y}_{i,t}|\Psi_i)} \frac{\mathscr{L}_t(\Psi_i)}{\mathscr{L}_t(\Psi_i^c)}\right)$$

$$= \min\left(1, \frac{p(\Psi_i^c)q(\Psi_i)}{p(\Psi_i)q(\Psi_i^c)} \prod_{t=1}^{T} \frac{\Pr(\mathbf{Y}_{i,t}|\Psi_i^c)}{\Pr(\mathbf{Y}_{i,t}|\Psi_i)}\right). \quad (16)$$

Although the relation between Equations (15) and (16) serves as an intuitive justification of the exchange algorithm, Liang et al. (2016) provide a formal proof of stationarity in the posterior distribution $p(\Psi_i|\mathbf{Y}_i)$.

Some may view the pseudo-likelihood (*PL*) proposed by Besag (1972) as a simple, viable alternative. The *PL* conveniently approximates the joint likelihood function by the product of full conditional distributions:

$$PL(\Psi_i) = \prod_{t=1}^{T} \prod_{k=1}^{K} \Pr\left(Y_{i,t,k} = 1|\Psi_i, \mathbf{Y}_{i,t,-k}\right)^{Y_{i,t,k}} \\ \cdot \Pr\left(Y_{i,t,k} = 0|\Psi_i, \mathbf{Y}_{i,t,-k}\right)^{(1-Y_{i,t,k})}, \quad (17)$$

where $\Pr(Y_{i,t,k} = 1|\Psi_i, \mathbf{Y}_{i,t,-k})$ is respondent i's conditional choice probability of item k in menu t defined in Equation (4) (see also Equation (8)). The conditional choice probabilities are in the form of binomial logits, and their product is easily computed even for extremely large menus. For the special case of $\Theta = \mathbf{0}$, but only for this special case, we have that the product of conditional choice probabilities in Equation (17) yields Equation (1).

In the presence of utility interactions, that is, with $\Theta \neq \mathbf{0}$, the *PL* is only large-sample consistent (e.g., Särkkä 1996 and Zhao and Joe 2005). Other authors report that the difference between PL and likelihood-based inference is only in efficiency (Gong and Samaniego 1981). However, the situation is somewhat more complicated in the context of hierarchical models. The successive conditioning on $\mathbf{Y}_{i,t,-k}$ inherent to the *PL* violates the assumption of independence between the distribution of conditioning arguments and the distribution of parameters across respondents. The resulting biases can be substantial (e.g., Manchanda et al. 2004 and Liu et al. 2007).

4.2. Hierarchical Prior

The standard hierarchical prior formulation, that is, a multivariate normal prior coupled with the likelihood implied by Equations (1) and (14) is well suited for parameters ($\beta_{i,1}, \ldots, \beta_{i,k}, \ldots, \beta_{i,K}$, and $\beta_{i,price}$) that characterize the attractiveness of individual items. It is less useful for inference about the utility-interaction parameters in Θ_i. The reason is heterogeneity in what respondents perceive to be perfect substitutes (or strong complements).

Recall that utility interactions $\theta_{i,k,k'}$ in Θ_i measure interactions in direct utility. For example, if respondent i receives no additional direct utility at all from item k' once the respondent chooses item k and vice versa, the corresponding parameter $\theta_{i,k,k'}$ is negative infinity for this respondent. However, if another respondent perceives horizontal differences between items k and k' such that choosing k and k' together becomes a possibility, $\theta_{k,k'}$ for this respondent is "infinitely" larger than for respondent i.

Figure 1. Mapping from $\mathbf{w}'_{k,k'}, \delta$ to θ Classes
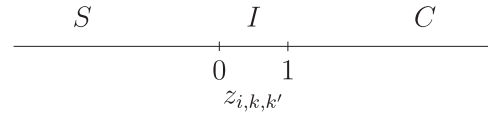
The standard approach to accommodating such extreme forms of heterogeneity is to use a hierarchical prior defined as a discrete mixture of distributions. In the following, we develop a hierarchical prior for Θ_i based on this idea. However, we pool information across individual elements $\theta_{i,k,k'}$ based on a prior understanding of the nature of the bivariate interactions. Pooling across individual elements $\theta_{i,k,k'}$ is necessary because of the limited amount of likelihood information to inform individual parameters $\theta_{i,k,k'}$ in Θ_i. The individual level likelihood of a respondent's choices across a (short) sequence of menus is maximized by setting all $\theta_{i,k,k'}$ that correspond to items that were never chosen together to negative infinity, for example. The implication that all these pairs are perfect substitutes to this respondent is likely to be wrong.

Our hierarchical prior classifies each element $\theta_{i,k,k'}$ for all $k = 1, \ldots, K$, $k' = k + 1, \ldots, K$, $i = 1, \ldots, N$ into one of three different "classes" based on an ordinal-probit regression. Equation (18) shows the "latent" linear equation underlying the ordinal-probit, and Figure 1 depicts how the latent dependent variable $z_{i,k,k'}$ in this regression maps into the distinct classes denoted S, I, and C based on fixed truncation points at zero and one. The class labels S, I, and C are intentionally chosen to relate to *s*ubstitution, *i*ndependence, and *c*omplemenarity. We revisit this point later.

$$z_{i,k,k'} = \mathbf{w}'_{k,k'}\delta + \epsilon_{i,k,k'}, \quad \epsilon_{i,k,k'} \sim N(0, \sigma_\epsilon^2). \quad (18)$$

Prior knowledge about $\theta_{i,k,k'}$ is encoded in the design vector $\mathbf{w}_{k,k'}$. If a particular bivariate interaction $\theta_{i,k,k'}$ is a priori expected to be similar to that between k and k'', that is, $\theta_{i,k,k'} \simeq \theta_{i,k,k''}$ for all $i = 1, \ldots, N$, for example, then $\mathbf{w}_{k,k'}$ is equal to $\mathbf{w}_{k,k''}$. However, the classification implied by the ordinal-probit model is probabilistic for all $\sigma_\epsilon^2 > 0$ and $-\infty < \mathbf{w}'_{k,k'}\delta < \infty$, such that the likelihood information in a respondent's choices may result in a posterior classification of a particular element $\theta_{i,k,k'}$ different from its (most likely) prior classification.

Conditional on classifications of individual $\theta_{k,k',i}$, we specify hierarchical normal, inverse gamma (*N-IG*) priors for the distribution of θ elements in each class:

$$\bar{\theta} \sim N(0, a^{-1}) \\ V_\theta \sim IG(v, s). \quad (19)$$

To compensate for the limited individual-level likelihood information and to regularize our inference

problem, we assume that θ elements, given classifications, are (relatively) tightly distributed around some location. For the S- and C-classes, we essentially fix V_θ equal to one but use a diffuse prior for $\bar{\theta}$. For the I-class $\bar{\theta}$ is fixed at zero and V_θ constrained to be practically zero.

To illustrate further, consider the following matrix of design vectors $\mathbf{w}'_{k,k'}$:

$$\mathbf{W} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

The columns correspond to coefficients $\boldsymbol{\delta}$ in Equation (18), in which the first coefficient is a constant and the following coefficients measure departures from that constant. For example, $\boldsymbol{\delta} = (.5, -5, 6)'$ implies that the θ parameter corresponding to line one in the preceding matrix and all other θ parameters with the same design vector a priori most likely connect two *i*ndependent items such that they should be zero (see Figure 1). For lines two and three in the matrix, these $\boldsymbol{\delta}$ values imply that the corresponding θ parameters are similar but different from zero. In principle, the likelihood determines if they are smaller (*s*ubstitution) or larger than zero (*c*omplementarity). However, depending on the prior knowledge about, thus, grouped θ elements, the prior expectation in one of the two directions will be strong.

For lines four and five in the matrix, these $\boldsymbol{\delta}$ values again imply that the corresponding θ parameters are similar, different from zero, and likely different from the θ parameters corresponding to lines two and three. Again, the likelihood determines if they are smaller (*s*ubstitution) or larger than zero (*c*omplementarity). And again, depending on the prior knowledge about, thus, grouped θ elements, the prior expectation in one of the two directions will be strong. We envision that, in most applications, prior information is rich enough to both establish expected similarities among θ parameters as well as an expectation about the directional departure from zero for at least some of the resulting groupings.

More formally, our hierarchical ordinal-probit model implies the following prior classification probabilities for θ parameters depending on the corresponding $\mathbf{w}$ vectors, and parameters $\boldsymbol{\delta}$ and σ_ϵ^2 in the hierarchical classification prior:

$$\Pr\left(\theta_{i,k,k'} \in S | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}, \sigma_\epsilon^2\right) = \int_{-\infty}^{0} N\left(z_{k,k'} | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}, \sigma_\epsilon^2\right) dz \quad (20)$$

$$\Pr\left(\theta_{i,k,k'} \in I | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}, \sigma_\epsilon^2\right) = \int_{0}^{1} N\left(z_{k,k'} | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}, \sigma_\epsilon^2\right) dz \quad (21)$$

$$\Pr\left(\theta_{i,k,k'} \in C | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}, \sigma_\epsilon^2\right) = \int_{1}^{+\infty} N\left(z_{k,k'} | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}, \sigma_\epsilon^2\right) dz. \quad (22)$$

The hierarchical prior on individual elements $\theta_{i,k,k'}$ established in Equations (18)–(22) has the following desirable properties: (i) it allows a collection of θ parameters believed to be similar a priori to take extreme values such that they can, for example, reflect essentially perfect substitution among a collection of items in a menu because (ii) it does not impose these extreme values for all respondents as the posterior classification of an element $\theta_{i,k,k'}$ may depart from its prior classification. Moreover, the posterior of the parameters in the hierarchical classification prior, that is, δ and σ_ϵ^2 in Equations (18) and (20)–(22) reveal if the prior grouping of θ parameters is supported by the data or not. For example, $\delta \neq 0$ coupled with small σ_ϵ^2 points to agreement between prior and posterior classifications. In summary, our adaptive prior anticipates concentrations of θ parameters around some negative value ($\bar{\theta}^{(S)}$) at zero ($\bar{\theta}^{(I)} \equiv 0$) and around some positive value ($\bar{\theta}^{(C)}$).

A theoretical drawback of the prior developed here is that it allows items to be classified as substitutes a posteriori even if the prior expectation is that they are complements and vice versa. In many applications, such reclassifications are regarded as a result of overfitting the data. In this case, it is useful to restrict the applicability of Equations (20)–(22) to a subset of θ parameters and to apply the following prior classification rules to θ parameters that are believed to connect, for example, substitutes, a priori:

$$\Pr\left(\theta_{i,k,k'} \in S | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}^S, 1\right) = \int_{-\infty}^{0} N\left(z_{k,k'} | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}^S, 1\right) dz \quad (23)$$

$$\Pr\left(\theta_{i,k,k'} \in I | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}^S, 1\right) = 1 - p\left(\theta_{i,k,k'} \in s | \mathbf{w}'_{k,k'}, \boldsymbol{\delta}^S, 1\right). \quad (24)$$

Equations (23) and (24) constrain the departure from the prior hypothesis of substitution between two items to independence and, thus, avoid the reclassification of prior substitutes into posterior complements deterministically. Obviously, the same approach can be used to avoid the reclassification of prior complements into posterior substitutes deterministically as well.

In our illustrative case study, we refer to this restricted prior as "MCMhet" and compare it to an MCM in which $\Theta_i \equiv \Theta$, that is, all utility interactions are assumed to be homogeneous. We refer to this model as "MCMhom."[12]

We defer the details of Bayesian inference based on the hierarchical priors discussed here to Online Appendix D. In this appendix, we develop approximate data augmentation for this model while maintaining the detailed balance conditions. Approximate data augmentation allows for automated construction of proposals that adapt to the current (and proposed) classifications of individual interaction parameters $\theta_{i,k,k'}$. This is necessary because of the large Metropolis–Hastings

steps that result from attempting to reclassify individual interaction parameters $\theta_{i,k,k'}$, and has the advantage that proposals need not be manually tuned.

5. Illustrative Case Study
5.1. Data
We illustrate our method using data from an MBCE featuring game consoles and console accessories. The experiment was conducted in 2013 by GfK for strictly academic purposes, however, mimicking typical client demands. Clients typically seek support for pricing decisions for new and existing products or product combinations (bundles) when commissioning MBCEs. As we show, data-based decision support is nontrivial because of complementary and substitutive relationships among game consoles and accessories in the MBCE, and the choice of model to connect the data to decision making matters substantively.

Respondents were recruited from an online panel and prescreened for market membership. If a prospective respondent had no console at home and was unlikely to purchase one within the next year, the experiment terminated after the screening questions. Overall, 575 respondents participated in the MBCE who were "in the market" according to the screening questions. Table 1 summarizes demographics of the sample used for estimation in our analysis. The median respondent in our sample is male, about 42 years old, lives in a three-person household, and has a high school degree. Overall, about 70% of respondents have at least one console among the set Xbox360, Xbox, PS3, PS2, Wii, or WiiU at home, whereby 41% own exactly one console, 19% two consoles, and 10% own more than two consoles. The remaining respondents either do not own game consoles (26%) or do not use any of the consoles they own any longer (4%). The average (median) console owner spends about 13 (6) hours per week playing video games with a standard deviation of about 20.

Table 2 summarizes the distribution of game genres played by respondents conditional on the type of console owned. For example, 69% of respondents who own a PS3 console use it for playing action games. Fitness games appear to be particularly popular among consumers who own a Nintendo console (Wii with 66% and WiiU with 62% compared with other consoles between 9% and 29%). On the other hand, Nintendo consoles are used less for playing adventure (Wii with 41% and WiiU with 33%) or racing games (Wii with 38% and WiiU with 29%). The differences between conditional game genre distributions in Table 2 are suggestive of horizontal differentiation between different consoles.

In our experiment, each individual choice menu was designed in the form of a webshop, in which respondents could put products into an electronic

Table 1. Sample Demographics

Variable	Categories	Percentage
Age	≤20	5.1
	21–40	41.5
	≥41	53.4
Gender	female	42.8
	male	57.2
HH size	1 person	15.4
	2 persons	32.9
	3 persons	24.7
	4 persons	21.0
	≥5 persons	6.0
HH income	<1,000	3.5
(monthly, in Euros)	1,000-2,000	15.7
	2,000-3,000	26.1
	3,000-4,000	23.6
	>4,000	12.8
	Missing	18.3
Education	Primary/secondary	12.4
	High school	57.2
	College/university	30.4

shopping basket by clicking on them. The shopping basket always clearly displayed the total expenditure, that is, the sum of the prices of individual items in the shopping basket to respondents. Respondents could explicitly indicate that they would not even purchase a single item from a particular menu using an additional no-choice option. Each respondent was exposed to 12 menus and clearly instructed to treat them independently. For each respondent, we exlcuded two menus from estimation for predictive validation. Each menu comprised the same 13 products, however, at varying prices.

Products presented in each menu were from three different brands, that is, Microsoft, Sony, and Nintendo. The design included the following game consoles and game console accessories: Xbox 360, Xbox One, PS3, PS4, Wii, Wii U, Xbox Kinect, Xbox Wheel, PS Eye Cam, PS Move, PS Wheel, Wii Wheel, and Wii Motion. The technical specifications of game consoles were constant across all menus, and respondents could instantly retrieve the information on technical details in each menu by hovering the mouse over a particular item. All consoles came with basic equipment, that is, were usable without purchasing additional items. The prices of individual items varied exogenously across different menus.

In our data, about 20% of respondents (122 out of 575) chose the outside option in every choice set. We exclude these respondents from our analyses as their data are void of information about relative preferences and substitution or complementarity among items in the menus. Thus, we rely on a sample of 453 respondents and 4,530 choices in total for estimation. Out of these 4,530 choices, at least one item is chosen on 3,403 occasions; 1,669 choices contain at least two

Table 2. Distribution of Game Genres Played Conditional on the Console Owned

	Action	Adventure	Racing	Sports	Fitness	Strategy	Other
PS3	0.69	0.59	0.60	0.48	0.18	0.30	0.00
PS2	0.51	0.54	0.54	0.43	0.09	0.22	0.09
Xbox	0.57	0.52	0.43	0.38	0.29	0.43	0.00
Xbox 360	0.68	0.60	0.59	0.51	0.19	0.30	0.00
Wii	0.33	0.41	0.38	0.50	0.66	0.16	0.03
WiiU	0.52	0.33	0.29	0.48	0.62	0.33	0.05

items from one menu, 613 choices at least three items, and finally 190 choices more than three items.

Table 3 details marginal and bivariate choice counts for the 13 products aggregated over respondents and repeated measurements. For example, PS4 was chosen 1,214 times in total, and PS4 and PS Wheel were chosen together 277 times, marginal with respect to other item choices that may have accompanied this pair. Inspection of the table reveals higher dependence between game consoles and accessories of the same brand compared with other pairs of products. Moreover, it is apparent that different consoles are sometimes chosen together and, therefore, are not uniformly viewed as perfect substitutes. Respondents, who chose two or more consoles together at least once spend more time on gaming than those who chose at most one console from a menu with means (medians) of 16.9 (9) and 12.02 (5) hours per week, respectively.

5.2. Models

In addition to our two MCM specifications—MCM with homogeneous Θ (MCMhom) and MCM with heterogeneous Θ with informative prior (MCMhet)—we include three benchmark models in our empirical model comparison: an independence model obtained from the MCM by setting all utility interactions to zero (Indep), the multivariate probit model (MvP), and an independence model with selected cross-price effects (IndepCPE). We note that all models included

in this comparison account for heterogeneity in preferences for individual items and the price parameter using a standard unrestricted multivariate normal hierarchical prior. The models differ only in the way for which dependence between items in a menu is accounted.

MCMhet uses the following two rules in its hierarchical prior specification: (i) utility interactions; that is, θ parameters that connect two consoles are a priori similar to each other and expected to be negative. As an exception, a particular interaction $\theta_{k,k',i}$ may equal to zero, but cannot be positive; (ii) utility interactions that connect a console to accessories of the same brand are a priori similar and expected to be positive. As an exception, a particular interaction $\theta_{k,k',i}$ may equal to zero but cannot be negative.

5.3. Results

Table 4 summarizes posterior means of alternative specific constants and price sensitivity at the hierarchical prior level. Posterior standard deviations (posterior SD) as measures of posterior uncertainty are reported in parentheses. The rank order of mean preferences across models is mainly consistent. For example, newer versions of the game consoles (Xbox One, PS4, and Wii U) deliver, on average, higher utility to respondents than older versions (Xbox 360, PS3, and Wii), and consoles deliver higher standalone utility than accessories. However, IndepCPE infers that some accessories deliver higher or about the

Table 3. Frequencies of (Pairwise) Choices

	Xbox 360	Xbox One	Xbox Kinect	Xbox Wheel	PS3	PS4	PS Eye	PS Move	PS Wheel	Wii	Wii U	Wii Wheel	Wii Motion
Xbox 360	290												
Xbox One	12	345											
Xbox Kinect	67	62	222										
Xbox Wheel	83	75	55	243									
PS3	35	9	10	20	395								
PS4	29	81	24	39	30	1214							
PS Eye	12	20	31	24	69	215	402						
PS Move	13	34	28	41	65	305	91	481					
PS Wheel	16	27	13	48	67	277	112	158	432				
Wii	42	14	28	32	34	43	26	17	20	538			
WiiU	14	42	26	44	6	90	28	37	31	41	452		
Wii Wheel	21	30	39	54	15	37	32	37	39	116	134	433	
Wii Motion	31	35	35	48	25	64	40	55	46	238	201	193	634

Table 4. Hierarchical Prior Means: Price Sensitivity and Alternative Specific Constants (Posterior SD)

	MCMhet	MCMhom	Indep	MvP	IndepCPE
Price	−0.34	−0.30	−0.30	−0.13	−0.67
	(0.02)	(0.02)	(0.02)	(0.01)	(0.04)
Xbox 360	−0.20	−0.66	−2.05	−1.34	4.70
	(0.32)	(0.31)	(0.38)	(0.19)	(1.26)
Xbox One	3.67	2.43	0.63	0.37	8.88
	(0.47)	(0.42)	(0.50)	(0.26)	(1.10)
Xbox Kinect	−3.33	−2.48	−3.39	−2.20	4.23
	(0.29)	(0.27)	(0.35)	(0.20)	(0.75)
Xbox Wheel	−4.39	−4.56	−4.77	−2.78	5.33
	(0.34)	(0.32)	(0.45)	(0.18)	(0.86)
PS3	0.10	0.26	−1.24	−0.83	1.89
	(0.35)	(0.28)	(0.34)	(0.16)	(0.80)
PS4	5.13	4.08	3.60	1.77	12.69
	(0.49)	(0.45)	(0.54)	(0.23)	(1.03)
PS Eye	−3.54	−3.08	−3.04	−1.95	5.81
	(0.25)	(0.20)	(0.23)	(0.13)	(0.73)
PS Move	−1.96	−1.95	−1.88	−1.23	12.02
	(0.26)	(0.25)	(0.26)	(0.14)	(1.03)
PS Wheel	−4.20	−4.47	−4.04	−2.44	6.81
	(0.25)	(0.29)	(0.35)	(0.19)	(0.75)
Wii	−0.59	−0.79	−1.97	−1.02	6.97
	(0.44)	(0.37)	(0.47)	(0.22)	(0.62)
WiiU	2.57	1.29	0.52	0.05	9.60
	(0.42)	(0.43)	(0.45)	(0.26)	(0.97)
Wii Wheel	−4.33	−3.57	−4.07	−2.24	5.40
	(0.30)	(0.33)	(0.37)	(0.17)	(0.64)
Wii Motion	−3.23	−2.99	−3.12	−1.86	6.03
	(0.25)	(0.25)	(0.29)	(0.15)	(0.59)

same utility as consoles. Compare, for example, the consoles PS3 and PS4 with mean utilities of 1.89 and 12.69 to the accessory PS Move with mean utility 12.02 and the console Wii to the accessory Wii Motion with mean utilities of 6.97 and 6.03 (column "IndepCPE" in Table 4).

The Indep model, that is, the version of the MCM that constrains all utility interactions to zero, infers all game consoles to be less attractive than when utility interactions are accounted for (compare Indep to MCMhet and MCMhom in Table 4). Because Indep has no way to account for substitution between game consoles, it cannot but suppress the attractiveness of individual game consoles to rationalize the observed data. Coefficients estimated by the MvP are absolutely smaller because of the difference in error normalization in the probit and logit models. For a summary of hierarchical prior standard deviations as measures of heterogeneity, see Table 5 in Online Appendix F.

For a more intuitive comparison across models at this level, consider the willingness-to-pay differential between PS3 and PS Move computed at the posterior mean of the hierarchical prior mean. MCMhet, MCMhom, Indep, and MvP imply willingness-to-pay differentials

of about 61, 74, 21, and 31 euros, respectively. The difference between Indep and the MCMs is related to the downward (upward) bias in the estimated preferences for consoles and accessories in Table 4. The difference between MvP and the MCMs—in line with our theoretical arguments—suggests that accounting for error correlation is insufficient to bring out the difference between the valuations of consoles and accessories. Finally, IndepCPE implies a willingness-to-pay differential between PS3 and PS Move of *minus* 405 euro, already suggesting that IndepCPE is misspecified in this context.[13]

Moreover, models that fail to capture substitutive and complementary relationships underestimate (overestimate) preference correlations between different consoles (consoles and corresponding accessories) in an attempt to rationalize the lack (prevalence) of joint choices. For example, MCMhet, Indep, and MvP infer population preference correlations between the two consoles Xbox One and PS4 of 0.78 (0.04), 0.56 (0.08), and 0.50 (0.07) and population preference correlations between the console Xbox 360 and the accessory Xbox Kinect of 0.19 (0.11), 0.53 (0.07), and 0.42 (0.08), respectively (see Tables 6–10 in Online Appendix F for the complete set of preference correlations from all models).

MCMhet relies on a hierarchical prior distribution for individual utility interactions (see Section 4.2). Table 5 summarizes the posterior of the hierarchical prior parameters that characterize the distribution of θ elements classified as connecting *i*ndependent (*I*), *s*ubstitutive (*S*), and *c*omplementary (*C*) items. Table 6 reports posterior means of hierarchical prior classification probabilities for θ elements into the three classes *I*, *S*, and *C* for MCMhet. Together Tables 5 and 6 strongly suggest that utility interactions are relevant for choices from the menus in our experiment. The posterior means of $\bar{\theta}^{(S)}$ and $\bar{\theta}^{(C)}$ are sizable, relative to alternative specific constants in Table 4 and well separated from $\bar{\theta}^{(I)} \equiv 0$ even after taking heterogeneity (V_θ) into account.

The posterior means of hierarchical prior classification probabilities in Table 6 vary by (soft) rule and in the expected directions. This supports the usefulness of these rules as a basis for probabilistically structuring the high-dimensional distribution of θ elements.

Table 5. Hierarchical Prior Means and Variance: Substitution and Complementarity—MCMhet (Posterior SD)

	I	*S*	*C*
Means ($\bar{\theta}$)	−0.01	−6.46	5.25
	(0.01)	(0.08)	(0.13)
Variance (V_θ)	0.09	1.05	1.01
	(0.00)	(0.05)	(0.04)

Table 6. Prior Classification Probabilities: Substitution and Complementarity—MCMhet

	Baseline	Two consoles	Console own accessory
Probability of I, %	75.63	24.27	78.38
Probability of S, %	23.32	75.73	0
Probability of C, %	1.05	0	21.62

For example, an element $\theta_{i,k,k'}$ connecting two consoles in MCMhet a priori belongs to the substitute class with a probability of about 76%. With a probability of about 24%, it a priori belongs to the independence class. This is consistent with the idea that a sizeable fraction of the population views consoles as horizontally differentiated in direct utility. Similarly, an element connecting a console to an accessory of the same brand a priori belongs to the complement class with a probability of about 22%. Table 6, thus, shows that respondents are heterogeneous in what they perceive to be (near) perfect substitutes or strong complements.

Table 7 reports demand interactions between game consoles across different models. For the heterogeneous versions of the MCM, we report posterior means of $E_i(\Theta_i)$, where E_i denotes the expectation across respondents. In parentheses, we report corresponding posterior standard deviations. As alternative measures for comparison, we report error correlations from the MvP as well as posterior means of average cross-price effects from IndepCPE. Because cross-price effects need not be symmetric, we report the effect from the item on the left to the item on the right in the first column of Table 7 in one column and the effect in the other direction in the next column.[14]

Both versions of the MCM infer strongly negative demand interactions among different consoles on average (see columns "MCMhet" and "MCMhom" in Table 7). Interestingly, assuming homogeneity in these interactions across respondents leads to relatively weaker interactions between consoles. Although the error correlations in the MvP reflect the sign of these negative interactions, only 7 out of 15 are credibly different from zero (see column "MvP" in Table 7). Finally, posterior means of average cross-price effects, reported in the last two columns of Table 7, lack face validity. All estimated cross-price effects between consoles are credibly negative on average. The negative cross-price effect from PS4 to Xbox One, for example, implies that Xbox One would never be chosen at an infinite price for PS4, suggesting that this model is misspecified. We further characterize the nature of this misspecification in Online Appendix E. Table 7, therefore, illustrates (i) the effect of accounting for heterogeneity in utility interactions and (ii) the decreased precision in how the MvP reflects utility interactions as

Table 7. Demand Interactions Between Game Consoles (Posterior SD)

	MCMhet	MCMhom	MvP	IndepCPE →	IndepCPE ←
Xbox 360–Xbox One	−4.91	−2.74	−0.03	—	—
	(0.18)	(0.13)	(0.05)	—	—
Xbox 360–PS3	−4.82	−1.10	−0.17	—	−0.41
	(0.14)	(0.11)	(0.07)	—	(0.15)
Xbox 360–PS4	−4.86	−1.08	−0.27	—	—
	(0.15)	(0.1)	(0.06)	—	—
Xbox 360–Wii	−4.79	−0.94	−0.02	—	−0.45
	(0.16)	(0.04)	(0.05)	—	(0.07)
Xbox 360–Wii U	−4.89	−1.23	−0.09	—	—
	(0.17)	(0.13)	(0.06)	—	—
Xbox One–PS3	−4.93	−0.97	−0.15	—	—
	(0.17)	(0.05)	(0.08)	—	—
Xbox One–PS4	−4.96	−1.72	−0.34	−0.14	−0.45
	(0.16)	(0.04)	(0.08)	(0.02)	(0.04)
Xbox One–Wii	−4.91	−0.95	−0.05	—	—
	(0.17)	(0.08)	(0.04)	—	—
Xbox One–WiiU	−4.87	−0.85	−0.06	—	—
	(0.15)	(0.08)	(0.08)	—	—
PS3–PS4	−5.08	−3.03	−0.36	−0.17	−0.34
	(0.13)	(0.05)	(0.05)	(0.03)	(0.04)
PS3–Wii	−4.83	−1.37	−0.24	—	−0.26
	(0.15)	(0.05)	(0.06)	—	(0.08)
PS3–WiiU	−4.97	−1.66	−0.36	—	—
	(0.13)	(0.13)	(0.08)	—	—
PS4–Wii	−4.79	−0.54	−0.12	—	—
	(0.16)	(0.05)	(0.08)	—	—
PS4–WiiU	−4.81	−0.91	0.00	−0.37	—
	(0.16)	(0.06)	(0.07)	(0.06)	—
Wii–WiiU	−4.99	−2.18	−0.15	−0.50	−0.57
	(0.15)	(0.11)	(0.05)	(0.10)	(0.03)

well as (iii) the failure of IndepCPE at even reflecting the sign of utility interactions.

In terms of interactions between consoles and accessories carrying the same brand, all models agree directionally (see Table 17 in Online Appendix F). Both versions of the MCM infer credibly positive demand interactions on average. Similarly, all error correlations in the MvP but one (PS3-PS Eye) as well as all estimated cross-price effects in IndepCPE are credibly positive and negative, respectively.

We next illustrate exemplary differences in inferred cross-price elasticities across models. Note that Indep and MvP imply zero cross-price elasticities by the definition of these models (see Equation (11) with $\theta = 0$ and Equation (12)). We, therefore, compare cross-price elasticities for MCMhet and IndepCPE only. We showcase the differences between MCMhet and IndepCPE looking at two illustrative respondents. Respondent 112 exhibits strong positive preferences for PS4 and PS accessories. Respondent 172 has positive preferences for consoles Xbox One, PS3, PS4, Wii, and

some of the accessories. Furthermore, respondent 172 never chooses two consoles together.

Elasticity computations are based on the price vector: 165 (Xbox 360), 383 (Xbox One), 117 (Xbox Kinect), 39 (Xbox Wheel), 165 (PS3), 325 (PS4), 39 (PS Eye), 117 (PS Move), 39 (PS Wheel), 165 (Wii), 299 (WiiU), 39 (Wii Wheel), and 39 (Wii Motion). These prices reflect average prices used in the experimental design. However, the price of PS4 is somewhat higher to stay in the elastic part of the choice-probability curves. We compute exemplary cross-price elasticities by forward simulating changes in market shares of products given a 1% increase in the price of PS4 as well as a 1% increase in the price of PS Wheel. Thus, reported elasticities relate to posterior changes in market shares that are relevant for decision making and not to posterior expectations of elasticities that may be meaningful in a model that directly parameterizes elasticities such as, for example, a constant elasticity model. For this reason, the reported elasticities do not have a distribution given the data (and the model).

Table 8 summarizes results. Own price elasticities are in bold. We see that IndepCPE counterintuitively infers negative cross-price elasticities between PS4 and PS3 (−0.065) and PS4 and Wii U (−0.212) in the case of respondent 112 and negative cross-price elasticities between PS4 and Xbox One (−0.092) and PS4 and Wii U (−0.044) in the case of respondent 172. Moreover, IndepCPE infers a negative cross-price elasticity between PS Wheel and PS Move for respondent 172 that is somewhat counterintuitive given that PS Wheel and PS Move are alternative means of interacting with PS-consoles.

We elaborate on the nature of the apparent misspecification of IndepCPE in Online Appendix E. In a nutshell, counterintuitive negative cross-price elasticities result from (similarly counterintuitive) negative cross-price effects that help IndepCPE rationalize no choice. In contrast, MCMhet estimates that increasing the price of PS4 increases the probability that our two illustrative respondents choose PS3 (or alternative consoles and their accessories) and decreases the probability of choosing PS-accessories.

We conclude that both IndepCPE and MvP fail at adequately representing substitution and complementarity. At the same time, our results demonstrate that demand interactions in the data are nontrivial. Moreover, respondents are heterogeneous in what they perceive to be near perfect substitutes and strong complements. Next, we investigate implications for predictive performance and optimal pricing.

5.4. Predictive Performance

Predictive performance is measured by hit rate (HR) and log predictive likelihood (LPL). We differentiate between marginal and menu HR. The former is defined as the proportion of correctly predicted choices of items or individual items in a menu (marginal HR), whereas the latter is defined as a proportion of correctly predicted choice combinations (menu HR). For example, if an individual chooses {A,B} out of a menu {A,B,C} and we predict {A-chosen, B-chosen, C-not chosen}, we declare it as three hits for marginal HR and as one hit for menu HR. However, if we predict, for example, {A-chosen, B-chosen, C-chosen}, we count this as two hits and one miss for marginal HR and as a miss for menu HR although the choice of {A,B} was predicted correctly. Because HR treats the case with predicted probability of, for example, 0.51 to be the same as 0.99, it is difficult to say which model is more

Table 8. Exemplary (Cross-) Price Elasticities

	Xbox 360	Xbox One	Xbox Kinect	Xbox Wheel	PS3	PS4	PS Eye	PS Move	PS Wheel	Wii	WiiU	Wii Wheel	Wii Motion
Respondent 112													
MCMhet ...													
PS4	0.020	0.038	0.027	0.010	0.049	**−0.061**	−0.010	−0.021	−0.046	0.021	0.065	0.002	0.014
PS Wheel	0.000	0.001	0.000	0.001	0.000	−0.001	0.000	0.000	**−0.014**	0.000	0.000	−0.001	−0.001
IndepCPE ...													
PS4	—	0.090	—	—	−0.065	**−0.205**	−0.117	−0.214	−0.090	—	−0.212	—	—
PS Wheel	—	—	—	—	0.019	—	—	0.019	**−0.035**	—	—	—	—
...													
Respondent 172													
MCMhet ...													
PS4	0.036	0.051	0.003	0.041	0.047	**−0.075**	−0.021	−0.014	−0.029	0.037	0.041	0.010	0.019
PS Wheel	0.000	0.000	0.000	0.000	−0.001	0.000	0.000	0.000	**−0.019**	0.000	0.000	0.000	−0.001
IndepCPE ...													
PS4	—	−0.092	—	—	0.113	**−0.129**	−0.233	−0.071	−0.092	—	−0.044	—	—
PS Wheel	—	—	—	—	−0.077	—	—	−0.152	**−0.057**	—	—	—	—
...													

Note. The numbers in bold indicate own price elasticities.

Table 9. Predictive Performance: Hit Rate (Log Predictive Likelihood)

Model	Marginal HR			Menu HR			
	1 & 0	1	0	Overall	None	One item	> 1 item
MCMhet	0.936 (−1,919)	0.585 (−1,182)	0.980 (−737)	0.575	0.876	0.512	0.468
MCMhom	0.938 (−1,921)	0.555 (−1,193)	0.985 (−728)	0.555	0.891	0.532	0.391
Indep	0.937 (−1,919)	0.555 (−1,204)	0.985 (−714)	0.531	0.905	0.520	0.332
MvP	0.937 (−1,928)	0.539 (−1,220)	0.986 (−709)	0.538	0.896	0.517	0.357
IndepCPE	0.935 (−2,376)	0.590 (−1,727)	0.977 (−650)	0.528	0.836	0.538	0.346

precise in its prediction. LPL overcomes this problem. LPL of observing the data can be expressed as

$$
\begin{aligned}
LPL = \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{t=1}^{T} & Y_{i,t,k} \log\left(\overline{\hat{p}}_{i,t,k}\right) \\
& + \left(1 - Y_{i,t,k}\right) \log\left(1 - \overline{\hat{p}}_{i,t,k}\right),
\end{aligned} \tag{25}
$$

where $\overline{\hat{p}}_{i,t,k}$ is the estimated *marginal* probability of choosing item k in menu t by individual i and $Y_{i,t,k}$ represents the realized choice of item k in menu t by individual i. To obtain $\overline{\hat{p}}_{i,t,k}$, we compute the marginal probability $\hat{p}_{i,t,k}$ at each draw of the posterior distribution of parameters by Gibbs sampling from the set of conditional probabilities that define the pseudo-likelihood (see Online Appendix C) and then take the average over posterior draws of parameters. Thus, we do not condition on holdout choices in any way making our predictions. The closer LPL is to zero, the better the predictive performance of the model.

Table 9 summarizes predictive performance of the models we compare in our two holdout menus. We do not find substantial differences between models in marginal HR (1&0), which is in the range between 93% and 94%. This can be explained by the prevalence of zeros in each choice; that is, only a relatively small number of items are chosen from each menu. Decomposing marginal HR into predictions of choices (marginal HR 1) and predictions of prevalent nonchoices (marginal HR 0), we find that the heterogenous MCM outperforms MvP in predicting marginal choices by about four percentage points. IndepCPE also performs well in terms of marginal HR 1 but is not on par with MCMhet in terms of LPL. Overall, MCMhet yields the most precise marginal predictions. As such, the LPL of MCMhet is −1,919, compared with, for example, the LPL of IndepCPE at −2,376 being the least precise model.

However, the crucial question in the analysis of data from MBCEs is how well a model predicts joint outcomes, that is, actually chosen item combinations.

Column (5) in Table 9 (menu HR, overall) shows that MvP, Indep, and IndepCPE are not competitive in terms of menu HR, trailing the performance of MCM that accounts for heterogeneity in utility interactions by about 4%–5%. An approximate standard error for the hit rates in this column is 0.017.

Interestingly, MCMhom, which estimates homogeneous Θ across individuals, performs relatively better too, suggesting that utility interactions in the MCM capture demand interdependencies present in our data better than the error correlations in the multivariate probit model. Unfortunately, we cannot compute the LPL for our menu predictions because of the difficulty of evaluating the likelihood that motivated the use of the exchange algorithm for inference earlier.[15]

Finally, we assess how well different models predict menu outcomes, depending on how many items were chosen in a menu. Columns (6)–(8) of the Table 9 summarize the results and compare menu HRs for none, one-item, and multiple-item choices, respectively. We find that, whereas the baseline models are competitive in predicting none and one-item choices, they are clearly inferior in terms of predicting multiple-item purchases. For instance, MCMhet predicts 46.8% of the multiple-item choices correctly, whereas MvP only 35.7%, resulting in a relative improvement of 30%. It is also apparent that accounting for heterogeneity in utility interactions in the MCM substantially improves predictions of multiple-item choices compared with an MCM with homogeneous Θ. An approximate standard error for the multiple item choice hit rates is 0.027.

5.5. Managerial Implications

Many firms allow customers to mass customize their products and services, and pricing of individual items is a major challenge. To meet this challenge, firms need to assess not only customers' preferences and price sensitivity but also customer-specific substitution and

complementarity relationships among items. We show next that pricing decisions substantially depend on if and how the demand model accounts for substitution and complementarity relationships among items. More specifically, we illustrate how optimal pricing patterns for complements and substitutes differ from those for independently valued items under different modeling assumptions.

We use the posterior obtained from the MCM with a structured account of heterogeneity in utility interactions (MCMhet) as well as that from baseline models, solve a given optimization problem, and compare the results. We find that accounting for substitution and complementarity among items in a menu results in substantially different optimal prices for products. The differences across models are in line with economic intuition about how accounting for substitution and complementarity should influence prices. We first consider optimal pricing for individual products that may or may not be purchased together, that is, the case of mass customization. Second, we illustrate implications for the pricing of preconceived bundles.

5.5.1. Optimal Pricing for Individual Products. For illustration, consider a monopolistic situation in which Sony optimizes prices for its products (PS3, PS4, PS Eye Cam, PS Move, and PS Wheel) given the competitor's products (Nintendo's Wii, WiiU, Wii Wheel, and Wii Motion and Microsoft's Xbox 360, Xbox One, Xbox Kinect, and Xbox Wheel) and prices in the market. Assuming marginal costs—set to zero without loss of generality—we define a five-dimensional grid of possible prices for Sony products ([69, 95, 99, 105, 169] for PS3, [199, 245, 249, 269, 299] for PS4, [19, 39, 99, 119, 139] for PS Eye Cam, [39, 59, 99, 109, 119] for PS Move, and [29, 39, 79, 99, 119] for PS Wheel, all prices in euros) and compute revenues for each price combination. We compute revenues for each individual in the sample based on posterior draws of individual parameters, β_i and Θ_i. We average over the draws of individual choice shares, compute individual revenues, then sum over all individuals, and finally find the price combination that maximizes the overall revenue.

Table 10 reports revenue-maximizing price vectors from different models. Indep and MvP, which do not account for utility interactions, price accessories too

Table 10. Revenue-Maximizing Prices

Model	PS3	PS4	Eye Cam	Move	Wheel
MCMhet	105	245	39	59	39
Indep	95	245	99	99	79
MvP	95	245	119	99	119
IndepCPE	69	199	139	119	29

Table 11. Marginal Shares at Revenue-Maximizing Prices

Model	Prices from	PS3	PS4	Eye Cam	Move	Wheel
MCMhet at	MCMhet	17.4	35.0	10.2	19.6	11.3
MCMhet at	Indep	17.8	33.1	5.0	13.0	7.0
MCMhet at	MvP	17.4	32.8	4.1	13.0	4.4
MCMhet at	IndepCPE	22.3	39.8	3.7	11.4	13.7
Indep at	Indep	17.6	33.7	3.8	9.8	5.3
MvP at	MvP	16.6	33.2	3.7	9.9	4.1
IndepCPE at	IndepCPE	21.6	43.6	4.1	21.0	16.5

high relative to the optimal prices for accessories from MCMhet. IndepCPE, which includes cross-price effects, yields even higher optimal prices for PS Eye Cam and PS Move. We also see that MCMhet yields relatively higher optimal prices for the two consoles PS3 and PS4.

The pattern in the optimal price vectors across different models illustrates an important managerial implication of substitution and complementarity among items in a menu. The complementarity between accessories and game consoles captured by MCMhet leads to lower prices for accessories. The relatively higher prices for consoles from this model—despite the complementarity to accessories—comes through strong substitution between PS3 and PS4. Whereas Indep and MvP predict that lower prices for PS3 and PS4 may likely result in a joint purchase, MCMhet correctly recognizes that this is an outlying event.

Table 11 reports market shares at revenue-maximizing prices. The first four lines in Table 11 show market shares computed using MCMhet, in turn conditioned on revenue-maximizing prices from the different models fitted. The following lines report shares computed using the models indicated in the first column at the optimal prices implied by that same model.

High optimal accessory prices recommended by the baseline models Indep and MvP (see Table 10) result in much lower market shares for accessories compared with those obtained from optimization based on MCMhet in Table 11. However, Indep and MvP tend to underestimate these shares (see lines "Indep at Indep" and "MvP at MvP"), presumably because these models do not account for the complementarity between consoles and accessories. Recall from Table 10 that IndepCPE suggests extremely high prices for Eye Cam and Move and relatively low prices for the two consoles. It vastly overestimates the shares of PS4 and Move at these prices (compare lines "MCMhet at IndepCPE" and "IndepCPE at IndepCPE" in Table 11). Finally, if a manager sets optimal prices suggested by IndepCPE, MvP, and Indep in this scenario, the loss in revenues relative to MCMhet is 5.65%, 3.53%, and 2.10%, respectively.

Next, we demonstrate how different models react to price changes in the market. We cut the prices of

Table 12. Reaction to Price Changes in the Market Δp

Model	PS3	PS4	Eye Cam	Move	Wheel
MCMhet	−10	−46	0	0	0
Indep	0	0	0	0	0
MvP	0	0	0	0	0
IndepCPE	−10	0	0	0	0

the competitive Microsoft and Nintendo items in half and reoptimize the prices for Sony products following the procedure described earlier. Table 12 summarizes the resulting price changes. For example, before the competitive price cut, MCMhet suggested a price of 245 for PS4 (see Table 10). After the competitive price cut, MCMhet suggests to lower this price by 46 to 199.

MCMhet implies to lower prices for PS game consoles and especially those of PS4. Accessory prices are not adjusted. IndepCPE reacts to the price changes too, but only by a small adjustment to the price of PS3. By definition, optimal prices are independent of competitors' prices in Indep that sets all demand interdependencies to zero, and thus, optimal prices stay the same. However, it may come as a surprise that the same holds for the MvP as explained analytically in Section 3.1 (see Equation (12)). Finally, if a manager sets optimal prices suggested by IndepCPE, MvP, and Indep under this competitive scenario, the loss in revenues relative to MCMhet amounts to 5.61%, 6.15%, and 5.13%, respectively.

Thus, if and in what way demand interdependencies are modeled matters substantially for the pricing of individual products consumers may or may not elect to purchase together. Given that the MCM, especially when coupled with a structured account of heterogeneity, is better supported by the data than Indep, MvP or IndepCPE, and the intuitive differences between optimal price vectors from different models, we conclude that an MCM coupled with a structured account of heterogeneity results in better pricing decisions in this and similar applications.[16] Next, we illustrate implications for optimal pricing of bundles in two stylized examples.

5.5.2. Optimal Bundle Pricing. For a variety of reasons, retailers and manufacturers offer preconceived bundles of products, sometimes without any possibility for mass customization (pure bundling) and sometimes in addition to offering products individually (mixed bundling). Here, we illustrate pricing implications in a pure bundling situation. An important advantage of the MCM is that—because it defines utilities for all possible bundles—we can predict choice probabilities over all sets of bundles that can be constructed from the products in the complete menu (cf. Jedidi et al. 2003).[17]

Note that both MvP and IndepCPE fail at this exercise. Both models leave choice probabilities for combinations of products that can only be purchased together unspecified, that is, *require* the specification of *product-specific* prices. Therefore, these models do not extrapolate to the action of asking one price for some combination of products when calibrated on menu choice data.[18] As a consequence, we compare MCMhet only to Indep, that is, the special case that assumes away utility interactions a priori, in this illustration.

We first investigate the choice between two "plausible" bundles in the sense that these two bundles reflect popular choice combinations in our data: "PS4 and PS Move" and "Wii and Wii Motion" (scenario 1 in Table 13). We again study optimal pricing from the perspective of a monopolist retailer who seeks to maximize revenues. Specifically, we determine the revenue-maximizing prices for the two bundles over a two-dimensional price grid with support at prices ending in "5" and "9."

Table 13 summarizes optimal prices from MCMhet and Indep with marginal shares in parentheses for this scenario in columns (2) and (3). Because MCMhet correctly recognizes that the bundles "PS4 and PS Move" and "Wii and Wii Motion" are relatively strong substitutes for most consumers, optimal prices are higher than those from Indep. At the same time, predicted shares are higher ("PS4 and PS Move") or just slightly smaller ("Wii and Wii Motion"). This is a consequence of the complementarity between PS4 and PS Move as well as between Wii and Wii Motion, which Indep ignores.

Scenario 2 replaces the second bundle, "Wii and Wii Motion," by a less plausible bundle featuring two consoles, "Xbox One and WiiU." Strong substitution between the bundles "PS4 and PS Move" and "Xbox One and WiiU" translates into a higher price for "PS4 and PS Move" than in scenario 1. Note that the pricing implications derived from Indep for the bundle "PS4 and PS Move" do not change because Indep ignores utility interactions (see columns (2) and (4) in Table 13). Also note the relatively low price for the bundle "Xbox One and WiiU" despite the fact that it comprises two of the latest consoles valued highly individually (column (5) in Table 13). The low price implied by MCMhet is a consequence of substitution between the two consoles making up this bundle. The even lower price implied by Indep is related to the high predicted market share (29.03%) for this bundle at a low price, which, in turn, is a function of ignoring substitution between XboxOne and WiiU comprising the bundle. Another reason for the low price from Indep is that it predicts a share of 11.46% choosing *both* bundles at these prices. In contrast, MCMhet puts the share of consumers choosing both bundles only at 1.9% in this scenario.

Table 13. Optimal Bundle Pricing (Marginal Shares)

	Scenario 1		Scenario 2	
	PS4 and PS Move	Wii and Wii Motion	PS4 and PS Move	XboxOne and WiiU
MCMhet	259	125	279	229
	(33.15)	(15.17)	(26.01)	(17.52)
Indep	239	109	239	209
	(30.79)	(16.40)	(30.79)	(29.03)

Our illustrations highlight the unique capability of the MCM to extrapolate from menu choice to the choice among bundles while taking substitutive and complementary relationships between products and between bundles comprising these products into account. Similarly, a firm benefits from the proposed model if it aims at forecasting the effect of price promotions on revenue or profit as well as when designing menus with and without bundles. More generally, the proposed methodology enhances managers' understanding of interrelated consumer demand and, thus, promises to help design more attractive offers.

6. Discussion

Studies that involve choices from menus have been gaining increased popularity. Given that consumers often face menu-like purchase situations and are commonly allowed to customize products and services based on menus, this is not surprising. Consumers' decisions in this case extend beyond the well-studied situation of choice among perfect substitutes.

In this paper, we propose the MCM as a utility-based framework to handle the combination of substitutes and complements likely to be present in many menus.

The parsimonious but flexible prior structure we develop captures extreme heterogeneity in two-way utility interactions. Such extreme heterogeneity in the form of "infinite differences" between consumers occur in populations in which some consumers perceive two items to be perfect substitutes and others view the same two items as horizontally differentiated.

We illustrate our methodology through an application to data from an industry-grade MBCE study eliciting demand for game consoles and accessories, designed by GfK for academic purposes. We find strong empirical evidence of substitution and complementarity between items in our experiment.

We show that approximations to utility interactions based on correlated errors, as in the multivariate probit model, or through cross-price effects result in inferior predictive fit, especially when predicting choices that comprise more than one menu item. Importantly, we document that the inferences from extant models result in counterintuitive pricing decisions and lower predicted revenues.

We trace the worse performance of extant models, that is, the multivariate probit and the binomial logit model that includes cross-price effects to the failure of generating important implications from the economics of substitution and complementarity. Specifically, we show analytically that the error correlations in the multivariate probit model can reflect substitution and complementarity between items in a menu. However, when the indirect utility of substitutes and complements changes, error correlations fail at generating the implied change in marginal shares of a target offering.

The binomial logit model that includes cross-price effects attempts to capture substitution through positive and complementarity through negative cross-price effects. However, substitution and complementarity exert their influence on individual choices even in the absence of price variation. And, in those instances in which price variation is sufficient for cross-price effects to pick up and reflect substitutive and complementary relationships, cross-price effects still fail at generating substitution between items whenever prices of substitutes equate their indirect utilities. Last but not least, the MCM is the only model that generalizes to the prediction of demand for arbitrary bundles, that is, fixed combinations of items (ex post) at some price even when calibrated on unrestricted choices from menus while taking complementarity and substitution into account.

On the technical side, we develop Bayesian inference for a hierarchical version of the MCM leveraging the recently proposed exchange algorithm in combination with perfect sampling of data from the MCM to overcome the problem of a computationally intractable normalizing constant in the likelihood. We also develop an auxiliary data augmentation scheme to automatically construct effective joint proposals of highly correlated parameters while maintaining the detailed balance conditions. Importantly, the proposed algorithm converges to the exact posterior implied by our model and priors.

Computationally, larger menus translate into more computational effort in the perfect sampler and when constructing proposal densities for updating parameters. The additional effort is linear in the size of the menu when constructing proposals and essentially linear

in the perfect sampler.[19] However, empirical identification tends to rely more on the hierarchical prior as the size of the menu increases with a fixed number of choices per respondent. As such, the proposed adaptive prior structure becomes even more important in larger menus.

Although we believe that the advantages of the MCM over extant models for choices from menus are apparent from our paper and that accounting for heterogeneity in both alternative specific utilities as well as utility interactions is useful, if not necessary in many empirical applications, there is certainly room for additional development. First, like all extant models in this context, the suggested modeling framework ignores income effects. Second, the MCM in its current implementation is limited to two-way utility interactions. Depending on the application, selected higher order interactions may also be important. However, the development of a prior that adaptively selects relevant, potentially heterogenous higher order interactions involves substantial theoretical and algorithmic development—given the enormous search space implied—that we leave to future research.

Endnotes

[1] We use the word "item" to indicate elements of the menu. In applications, these items could be attributes of the products, parts of the products that work together, or complete products themselves.

[2] Note that, when applied to a market, pure vertical differentiation implies ordinal homogeneity in direct utility and horizontal differentiation different preference orderings for different consumers. When applied at the individual level, vertical differentiation, together with flat direct utility after the first unit, implies perfect substitution, that is, discrete choice among perfect substitutes. Horizontal differentiation at the individual level means that a single consumer receives different types of utility from different items as if they were in different categories such that items do not act as substitutes, and a joint purchase can be utility maximizing.

[3] Note that even small approximation biases at the individual level that are likely in the "double-MH" approach proposed by Liang (2010) or the adaptive exchange algorithm proposed by Liang et al. (2016) can aggregate to substantial overall biases in the setting of a hierarchical model that accounts for unobserved heterogeneity.

[4] Conditional on parameters Ψ_i, this is just a normalizing constant that can be dropped when sampling data from the model defined in Equation (1). However, because this normalizing constant depends on parameters, it cannot be ignored for estimation purposes.

[5] The superscripted time here is in arbitrary units reflecting human processing epochs. One might imagine the entire decision process to last a fraction of a second or many minutes.

[6] A menu with K items translates into $2^K = J$ possible choices, including the choice of the outside good. Normalizing the utility of the outside good to zero for identification and realizing that the K possible single-item choices from the menu do not involve utility interactions, we arrive at $2^K - K - 1$ interaction parameters.

[7] More formally, we then have that $\Pr(Y_A = 1) = \Pr(Y_A = 1, Y_B = 0) + \Pr(Y_A = 1, Y_B = 1) = \Pr(Y_A = 1, Y_B = 0) + 0 = \Pr(Y_A = 1, Y_B = 0)$; that is,

the choice of one item precludes the choice of the other because of perfect substitution. In contrast, when θ_{AB} increases to plus infinity, choices degenerate to choosing only and always the bundle $\{A, B\}$.

[8] Also note that including interaction terms in the utility functions of an MvP destroys the coherency between the collection of (binomial probit) item-choice probabilities and the implied probability distribution over bundles. In other words, "cycling through" the collection of (binomial probit) item-choice probabilities leads to a joint distribution of bundles in which conditional item choice probabilities *differ* from the data-generating collection of (binomial probit) item-choice probabilities.

[9] Recall that this is the denominator of the MNL-likelihood in Equation (1) that depends on unknown parameters and, thus, cannot be simply dropped from the MH-acceptance probability. Intuitively, dropping this normalization in estimation would allow for parameters that appear to explain well what is chosen, that is, yield large values for the numerator in Equation (1) but fail to motivate what was not chosen from a particular menu.

[10] We explain how to automatically construct good proposals in Online Appendix D, step 6b.

[11] For the MCMC defined in Equation (15) to exactly correspond to the likelihood of our model, the auxiliary $Y_{i,t}^c$ needs to be sampled independently from $\Pr(Y_{i,t}^c | \Psi_i^c)$. "Gibbsing through" the corresponding set of conditional distributions defined in Equation (4) converges to providing samples from $\Pr(Y_{i,t}^c | \Psi_i^c)$. However, it is not clear a priori for how long to run the Gibbs chain before this happens, which motivates the use of the perfect sampler here.

[12] Kwak et al. (2015) propose to structure Θ based on prior assumptions about homogeneity in perfect substitution.

[13] A one price unit difference in the design matrix corresponds to 10 euros. When computing the willingness-to-pay differential under IndepCPE, one needs to consider the cross-price effect of –0.42 from PS3 to PS Move (see Table 17 in Online Appendix F).

[14] For the complete set of demand interactions and corresponding measures of heterogeneity (where applicable) across different models, see Tables 11–16 in Online Appendix F. Cross-price effects to estimate in IndepCPE were chosen following the MBC industry guidelines (Orme 2013). These guidelines suggest to retain cross-price effects based on the statistical significance of cross-price dependencies established in aggregate count analyses.

[15] When simulating menu choices to estimate probabilities for the $2^{13} = 8,192$ potential choices from each menu as relative frequencies, there is always a chance that we obtain a probability of zero for a particular observed response because of the numerical constraints of simulation.

[16] Anecdotally, observed market prices are directionally consistent with the prices derived from MCMhet in that prices of game consoles are proportionally larger than those of accessories.

[17] In addition, the MCM implies degrees of substitution and complementarity between *bundles* as functions of the utility interactions between *products* defining the bundles. This compares favorably to defining the choice between bundles as a choice among perfect substitutes a priori.

[18] One can certainly fit IndepCPE, MvP, or even a multinomial probit model to choices among preconceived bundles but is then constrained to modeling this (smaller) fixed set of preconceived bundles.

[19] MvP is about equally computationally intensive because of the augmentation of dependently distributed latent utilities. Indep and IndepCPE are computationally light because the individual level likelihood is a product of conditionally independent binomial logits.

References

Besag JE (1972) Nearest-neighbour systems and the auto-logistic model for binary data. *J. Roy. Statist. Soc. B* 34(1):75–83.

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. B* 36(2):192–236.

Blattberg RC, Wisniewski KJ (1989) Price-induced patterns of competition. *Marketing Sci.* 8(4):291–309.

Carpenter GS, Cooper LG, Hanssens DM, Midgley DF (1988) Modeling asymmetric competition. *Marketing Sci.* 7(4):393–412.

Chung J, Rao VR (2003) A general choice model for bundles with multiple-category products: Application to market segmentation and optimal pricing for bundles. *J. Marketing Res.* 40(2):115–130.

Farquhar PH, Rao VR (1976) A balance model for evaluating subsets of multiattributed items. *Management Sci.* 22(5):528–539.

Gentzkow M (2007) Valuing new goods in a model with complementarity: Online newspapers. *Amer. Econom. Rev.* 97(3):713–744.

Gong G, Samaniego FJ (1981) Pseudo maximum likelihood estimation: Theory and applications. *Ann. Statist.* 9(4):861–869.

Hoch SJ, Kim B-D, Montgomery AL, Rossi PE (1995) Determinants of store-level price elasticity. *J. Marketing Res.* 32(1):17–29.

Jedidi K, Jagpal S, Manchanda P (2003) Measuring heterogeneous reservation prices for product bundles. *Marketing Sci.* 22(1):107–130.

Kamakura WA, Kwak K (2012) Menu-choice modeling. Working paper, Rice University, Houston, TX.

Kwak K, Duvvuri SD, Russell GJ (2015) An analysis of assortment choice in grocery retailing. *J. Retailing* 91(1):19–33.

Liang F (2010) A double Metropolis-Hastings sampler for spatial models with intractable normalizing constants. *J. Statist. Comput. Simulation* 80(9):1007–1022.

Liang F, Jin IH, Song Q, Liu JS (2016) An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants. *J. Amer. Statist. Assoc.* 111(513):377–393.

Liechty J, Ramaswamy V, Cohen SH (2001) Choice menus for mass customization: An experimental approach for analyzing customer demand with an application to a web-based information service. *J. Marketing Res.* 38(2):183–196.

Liu Q, Otter T, Allenby GM (2007) Investigating endogeneity bias in marketing. *Marketing Sci.* 26(5):642–650.

Manchanda P, Rossi PE, Chintagunta PK (2004) Response modeling with nonrandom marketing-mix variables. *J. Marketing Res.* 41(4):467–478.

Möller J, Pettitt AN, Reeves R, Berthelsen KK (2006) An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* 93(2):451–458.

Murray I, Ghahramani Z, MacKay D (2006) MCMC for doubly intractable distributions. *Proc. 22nd Annual Conf. Uncertainty Artificial Intelligence, Arlington, Virginia.*

Orme BK (2010) Menu-based choice modeling using traditional tools. *Sawtooth Software Conf. Proc., Sequim, WA,* 85–110.

Orme BK (2013) *MBC v1-Software for Menu-Based Choice Analysis* (Sawtooth Software, Orem, UT).

Propp JG, Wilson DB (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* 9(1–2):223–252.

Russell G, Petersen A (2000) Analysis of cross category dependence in market basket selection. *J. Retailing* 76(3):367–392.

Särkkä A (1996) Outlying observations and their influence on maximum pseudo-likelihood estimates of Gibbs point processes. Working paper, University of Gothenburg, Gothenburg, Sweden.

Wedel M, Zhang J (2004) Analyzing brand competition across subcategories. *J. Marketing Res.* 41(4):448–456.

Zhao Y, Joe H (2005) Composite likelihood estimation in multivariate data analysis. *Canadian J. Statist.* 33(3):335–356.