

Modeling Choice Endogeneity in Consumer Experiments

Fred M. Feinberg

University of Michigan Business School
701 Tappan Street, Ann Arbor, MI 48109
(734) 764-4711, FAX: (734) 764-3240
feinf@umich.edu

Christina L. Brown

University of Michigan Business School
701 Tappan Street, Ann Arbor, MI 48109
(734) 764-4717, FAX: (734) 764-3240
clbrown@umich.edu

Yuanping Ying

University of Michigan Business School
701 Tappan Street, Ann Arbor, MI 48109
(734) 615-8136, FAX: (734) 764-3240
yingyp@umich.edu

Under Review at *Marketing Science*. Please do not quote or distribute. All comments welcome.

The authors wish to thank Greg Allenby, Clint Cummins, Terry Elrod, Gabor Kezdi, Michel Wedel, Carolyn Yoon and Jie Zhang for their assistance and suggestions.

Modeling Choice Endogeneity in Consumer Experiments

ABSTRACT

A common goal in consumer behavior research is to understand how one evaluates products filtered through some type of screening or selection process. Typical examples include post-choice satisfaction ratings, certain “free recall” tasks, or the development of consideration sets followed by brand choice. In such situations, what the researcher observes is contingent not only on those alternatives being evaluated, but the process by which they have become available, creating a selection bias. That is, there may be pronounced *choice endogeneity*. When substantial endogeneity is present, gauging effect strengths through standard methods can introduce demonstrable biases. Models accounting for endogeneity are familiar in the economics literature (notably Heckman 1979), and are becoming increasingly important in quantitative marketing; however, the issue is largely absent from the consumer behavior literature. Moreover, certain formal aspects of standard endogeneity models limit their range of use in experimental studies of consumer judgment and decision making.

In this paper, we develop a class of models applicable to a broad range of decision problems in consumer behavior, and provide a set of tools for their estimation. In particular, we consider situations in which a polychotomous selection process limits which items offer up subsequent information. We illustrate the method by analyzing an experiment in which subjects choose among, and then evaluate, a frequently-purchased consumer good, as well as data first examined by Ratner, Kahn and Kahneman (1999). Results indicate not only strong endogeneity, but that traditional specifications of the choice process – independent binary selection, no endogeneity, and interval-scaled predictions – can lead to markedly different interpretations of variable effects. Response heterogeneity in and of itself does not alter our substantive results.

KEYWORDS: Choice Models, Consumer Behavior, Decision-Making, Econometric Models, Sample Selection, Endogeneity, Heckman Model, Markov chain Monte Carlo, Hierarchical Bayes

1 Introduction

Sample selection bias is a common problem in academic disciplines characterized by field research – labor economics and sociology, for example – in which it is not often possible to guarantee a random sample of the population of interest. In these fields, researchers have come to rely on statistical models designed to correct for the effects of self-selection and endogeneity. There is a long history of such models in economics, dating back to the classic papers of Tobin (1958) and Heckman (1976, 1979), and recently there has been a surge of interest in the topic among quantitative researchers in marketing, starting with the pioneering work of Villas-Boas and Winer (1998) (see Draganska and Jain 2003 for a comprehensive review).

Endogeneity has recently become a concern in a wide variety of marketing-related contexts: consumer choice dynamics under partial observability (Erdem, Imai and Keane 2003), share response in contiguous areas (Bronnenberg and Mahajan 2001), competitive pricing (Besanko, Gupta and Jain 1998) and/or advertising (Chintagunta, Kadiyali and Vilcassim 1999), as well as promotional (Boatwright, Dhar and Rossi 2003) and retail pricing effects (Huang, Hardie and Putsis 2001). Chintagunta (2001) found that ignoring endogeneity can even have a greater impact on estimated price elasticities than ignoring the effects of heterogeneity. It has been long known that field data require careful treatment to circumvent potential endogeneity effects, a central motivator of the “structural economic modeling” approach (e.g., Berry, Levinsohn and Pakes 1995; Reiss and Wolak 2002).

By contrast, corresponding demonstrations and methods suitable for studies in experimental consumer research are very much lacking. Because they often have the luxury of randomly assigning subjects to experimental conditions, laboratory researchers have been considered largely exempt from the need to be concerned with problems of selection and endogeneity. Degeratu, Rangaswamy, and Wu (2001) suggest that, to study internet purchase patterns, one “would ideally need to conduct a randomized experiment in which some people are assigned to shop online and some are assigned to shop offline.” They underscore the endogenous nature of field data, contrasting it with the controlled experimental settings typical to behavioral researchers. If people were randomly exposed to products (without giving thought to which they might like best), or were somehow randomly assigned to try some set of alternatives, product choice would be exogenous: literally outside their control. Random assignment in experimental settings is thus supposed to control for endogeneity. While in many cases it does, in others it does not.

In this paper, we show that choice endogeneity should be considered even in well-designed experimental

settings, and that selection or screening processes are frequently a critical, if unheralded, part of experimental research in marketing. For example, consider the following situations:

- A consumer chooses among alternatives, then indicates satisfaction with his/her choice, either on a continuous scale (Ratner, Kahn, and Kahneman 1999), or on a more limited or discrete one, such as a 1-to-7 scale (Brown 2003, Shiv and Huber 2000, Zhang and Fitzsimons 1999);
- A consumer chooses among alternatives, then makes a follow-up choice (Iyengar and Lepper 2000), or decides whether to switch to another brand (Fitzsimons 2000); or a consumer chooses what product information to examine, then chooses a brand (Braun 1999);
- A consumer develops a consideration set then makes a final choice (Haübl and Trifts 2000); or a consumer makes a sequences of choices reducing the set of alternatives by stages until only one remains (Tversky 1972, Levin et al. 1991, 1998).

These are commonly encountered experimental situations in which selection bias may be relevant, because only certain items ‘survive’ the selection to be observed at a later stage: we observe a post-choice satisfaction rating for Brand X only if it is the brand actually selected; we can observe a switch away from Brand Y only if Brand Y has indeed been chosen in the first place; we can observe choice of Brand Z only if Brand Z made it into the consideration set. Other consumer situations that have received less experimental attention may involve selection phases, such as rank-ordering a list of alternatives then choosing from the ranked list (as with web-based intelligent agents). In short, results stemming from later stages of such processes are of great interest to behavioral researchers, yet the earlier stages are rarely systematically accounted for.

One encounters three types of difficulties in applying the classic Heckman-type specification, consisting of a binary selection submodel and an OLS prediction submodel, to the consumer behavior examples above. First, the binary selection submodel assumes that each item in a choice set is considered *independently* of others. This is seldom the case when studying individual decision making: for example, in making a single choice among two alternatives, the choice of one eliminates the possibility of choosing the second, no matter how appealing it may be. Secondly, the OLS prediction submodel is predicated on a continuous, unbounded and linearly-related dependent variable. Evaluative outcomes in psychology and consumer research are generally measured on a discrete, limited scale that possesses only ordinal properties. Finally, selectivity models supported in standard software rarely allow for heterogeneity in consumer sensitivities, and never in whether selectivity itself differs across subjects.

Our main goals in this paper, therefore, are: to develop models that, while similar in spirit, are more directly applicable to the problems typically encountered in marketing, and other disciplines where choices and evaluative processes are commonly studied; to provide tools for their estimation; and to present empirical evidence for their importance using laboratory choice data. We take into account two processes: that by which items are valued or chosen (what we will call “selection”) and that by which items are evaluated (“prediction”).

We start by briefly reviewing the standard selectivity (Heckman type) model, cataloguing its strengths and weaknesses as a tool for analyzing consumer decisions. This will lead to extensions in both submodels, and we show how the resulting class of models can be readily estimated, both classically and through more general Bayesian techniques. The models’ relative implications are explored by analyzing experimental data on choice and post-choice satisfaction for a frequently-purchased consumer good, as well as data first examined in Ratner, Kahn and Kahneman (1999). Results not only indicate strong endogeneity, but also that more “traditional” specifications of the process (independent binary selection, no endogeneity, and interval-scaled predictions) lead to substantially different interpretations of variable effects, though response heterogeneity, intriguingly, does not. We conclude by illustrating how the same basic modeling and estimation approach readily applies to a number of alternative specifications useful for experimental applications.

1.1 Endogeneity and the Heckman Framework

The standard Heckman (1979) selectivity model is typically given by a two-stage system of the following type:

$$Y_s = X_s\beta_s + \epsilon_s, \tag{1}$$

$$Y_p = X_p\beta_p + \sigma\epsilon_p \text{ if } Y_s > 0, \tag{2}$$

$$(\epsilon_s, \epsilon_p) \sim N[0, 0, 1, 1, \rho]. \tag{3}$$

The (second-stage) prediction variable, Y_p , is observed only in cases where the (first-stage) selection variable, Y_s , is positive. Each equation has its own set of regressors, $\{X_s, X_p\}$, so that each could be readily estimated separately were their errors uncorrelated. The full system (1)-(3) is ordinarily estimated by maximum likelihood techniques, though early lack of computation power spawned an extensive literature on two-step estimation approaches (Nelson 1984); Lee (1983) and Winship and Mare (1992) provide numerous examples, as well as discussions of their shortcomings. Hay (1980) and Dubin and McFadden (1984) introduced similar selectivity models relying on two-step estimation techniques. When endogeneity, as measured by ρ ,

is substantial, the two-step approach can be quite inaccurate (Puhani, 2000), and so we use full information maximum likelihood (Nawata 1994) and modern MCMC methods instead.

It is important to highlight the role of ρ , which ‘picks up’ the correlation between the stochastic components of the selection and prediction submodels. Blundell and Powell (2003) suggest that “the analysis of data with endogenous regressors ... is arguably the main contribution of econometrics to statistical science.” They further point out that while “endogeneity can arise from a number of different sources, including mis-measured regressors, sample selection, heterogeneous treatment effects, correlated random effects in panel data, etc.,” at its core, the endogeneity problem makes itself known through correlations in observed or, in the case of latent variables, unobserved errors. The Heckman model reduces the issue to measuring a single error correlation, ρ , and doesn’t aim to directly pinpoint the source of that correlation, for example (as per Blundell and Powell), heterogeneity, omitted or mismeasured variables. Much literature attempts to address these other sources of misspecification, particularly in limited dependent variable models, and we direct readers to the recent paper of Honore and Hu (2002). In empirical applications using laboratory data, it is important that researchers explicitly check whether supposed endogenous effects – as captured by error correlations such as ρ in (3) – are robust to included exogenous regressors and coefficient heterogeneity. In the two data sets we examine herein, findings of endogeneity are indeed robust in these ways, though this cannot be presumed in general.

For all its elegance and successful application throughout macroeconomics, the system given by (1)-(3), has a number of limitations when applied in other domains. Two are directly relevant for the present study: a binary selection mechanism, and an interval-scaled dependent variable in the prediction equation. We consider these in turn.

1.1.1 Polychotomous vs. Binary Selection Submodels

The nature of the binary Selection submodel (1) dictates that each item entered into the selection equation is considered independently of all other items. This is justified when each of a set of items is considered on its own merits, as when one offers admission to a graduate program for any and all students surpassing a certain weighted index of standardized exam scores, past grades, and the like (Dawes 1979). Students are therefore never *directly* compared with one another, so that the qualities of one student might render another less likely to, literally, ‘make the cut’, at least once the cut is set (Feinberg and Huber 1996). In most marketing contexts, however, items do compete for inclusion. Consider choosing an entrée from a menu. Whether the restaurant is of high quality (many of the entrées are appealing) or poor (few or none are appealing), we do

not choose multiple items in the first case or zero in the second, but one in each. In terms of the selection submodel, then, we choose exactly one item from a given set, and all that will matter is the *between-item* comparison. Such a mechanism is foundational in brand choice models, which seek to explain *which* of a set of *competing* brands will be chosen, given that a *single choice* is observed.

We therefore consider a ‘polychotomous’ selection submodel appropriate to marketing contexts in which items *compete* for choice. That is, selection operates through a multinomial probit (discrete choice) model, as opposed to the (independent) binary probit mechanism of (1). While these may appear superficially similar, we will find that using a binary selection mechanism when a polychotomous one is called for can yield markedly different conclusions regarding effects strengths.

1.1.2 Continuous Interval-Scaled Prediction Submodel

The second-stage Heckman Prediction submodel (2) is based on an OLS framework, suited to an interval-scaled dependent variable. More common in marketing research, however, is a discrete, ordered scale with a limited number of points (such as brand attitude measured on a 1-to-7 scale). Treating such dependent variables as though they were interval-scaled creates a number of problems, for example, ceiling effects and out-of-range predictions. Interval scaling also assumes that “a 6 vs. a 5” and “a 7 vs. a 6” represent identical increases in consumer attitude. Presuming that each covariate comprising X_p produces identical unit changes throughout its range is an extraordinarily strong assumption, one which we shall empirically test against. We thus extend (2) to account for ordinal predictions, which we take to be generated by an underlying continuous reference variable, subject to cutoffs (as defined subsequently). By setting these cutoffs to evenly-spaced values on the unit interval, we can directly test for underlying response linearity, while still accommodating an ordinal observable.

Both submodels have been treated extensively in the econometric literature (Maddala 1983, Amemiya 1985, Agresti 1990). Our purpose is to examine them in the case where observation in the second stage is contingent on the outcome of the first, and when their errors cannot be presumed independent.

2 Data and Notation

To simplify exposition, it is helpful to refer to data for two specific subjects, who are faced with choice sets that may differ in composition and/or size:

<i>Respondent Number</i>	<i>Number Choices</i>	<i>Selection : Polychotomous</i>	<i>Prediction : Ordinal</i>	<i>Selection Covariates</i>	<i>Prediction Covariates</i>
r	k_r	Y_s	Y_p	X_s	X_p
1	2	1	10	$X_{s,1,[1:2]}$	$X_{p,1,[1:2]}$
1	2	0	–	$X_{s,1,[2:2]}$	$X_{p,1,[2:2]}$
2	3	1	9	$X_{s,2,[1:3]}$	$X_{p,2,[1:3]}$
2	3	0	–	$X_{s,2,[2:3]}$	$X_{p,2,[2:3]}$
2	3	0	–	$X_{s,2,[3:3]}$	$X_{p,2,[3:3]}$

We denote selection and prediction estimates as $U_{s,r,[i:k_r]} = X_{s,r,[i:k_r]}\beta_s$ and $U_{p,r,[i:k_r]} = X_{p,r,[i:k_r]}\beta_p$. The r subscript can be suppressed where clarity is not sacrificed, and, for simplicity, we number the alternative chosen (i.e., selected) by each subject as $[1 : k_r]$ or simply as 1. We therefore observe second-stage values only for $i = 1$, so that we refer unambiguously to $U_{p,[1:k_r]}$ or $U_{p,1}$.

2.1 Full Model: Polychotomous Selection, Ordinal Prediction and Endogeneity

As in the standard Heckman model (1)-(3), endogeneity is accounted for by considering joint error draws from a standard bivariate normal distribution, $(\epsilon_s, \epsilon_p) \sim N[0, 0, 1, 1, \rho]$.¹ The dependent measure in the prediction submodel, Y_p , is ordinal, and takes on consecutive integral values. This is modeled in the standard manner, as probabilities of falling between cutoffs, $\{\mu_j\}$, which are estimated. Thus, the joint probability for a particular observation is:²

$$P \left[U_{s,[1:k]} + \epsilon_{s,[1:k]} > \{U_{s,[i:k]} + \epsilon_{s,[i:k]}\}_{i>1} \text{ and } \mu_{j-1} < U_{p,[1:k]} + \sigma\epsilon_{p,[1:k]} \leq \mu_j \right], \quad (4)$$

where σ is an estimated dispersion parameter and the observed categorical response is j . In practice, for identification purposes, we estimate only a subset of these, setting the upper cutoff for the smallest observation in the data to zero and the lower cutoff for the largest to one. Thus, for c ordinal response categories, we estimate $c - 3$ cutoffs in $(0, 1)$, as well as the dispersion parameter σ , with the convention that $\mu_0 = -\infty, \mu_1 = 0 < \mu_2 < \dots < \mu_{c-1} = 1, \mu_c = \infty$.³

If $\{\epsilon_{s,[i:k]}\}$ are multinormal with zero mean and identity covariance matrix, (4) can be readily evaluated by integrating across $\epsilon_{s,[1:k]}$; because its only correlation is with $\epsilon_{p,[1:k]}$, the likelihood contribution is:

$$\int_{\theta \in \mathbb{R}} \phi(\theta) \left(\prod_{i>1} \Phi[\theta - (U_{s,[i:k]} - U_{s,[1:k]})] \right) \Delta_j \left[\Phi \left[\frac{(\mu_j - U_{p,[1:k]}) - \sigma\rho\theta}{\sigma\bar{\rho}} \right] \right] d\theta, \quad (5)$$

¹It is possible to impose a hierarchical specification on ρ itself, or to allow it its own (exogenous) covariates. We will explore the former possibility empirically.

²For conciseness, we use $x \geq \{y_i\}$ to mean $x \geq \max_i y_i$.

³This scaling for cutoffs differs slightly from the standard one, which sets $\sigma = 1$ and leaves μ_{c-1} free. We avoid this scaling because, all else equal, fixing σ to unity allows the degree of error in the prediction submodel to determine the overall scale for not only the cutoff values, but for the elements of β_p , so that prediction submodels with smaller error will appear to have larger effects coefficients. Setting the range of estimated cutoffs to $(0, 1)$ prevents this. Substantively, the two scalings are interchangeable, differing only by a constant multiple.

where Δ_j is the difference operator, $\Delta_j [f(x_j)] = f(x_j) - f(x_{j-1})$. It is straightforward to estimate the parameters implied by (5) using quadrature, simulated likelihood or other (classical) methods. If $\{\epsilon_{s,[i:k]}\}$ are multinormal with arbitrary to-be-estimated (positive definite) covariance matrix, estimation is complicated substantially, and one can appeal to Bayesian (MCMC) methods. In the Appendix, we show that it is possible to choose a scaling for this estimated covariance matrix which preserves the joint distribution $(\epsilon_s, \epsilon_p) \sim N[0, 0, 1, 1, \rho]$, and further, that this ‘sets the scale’ of the estimated effects coefficients $\{\beta_s, \beta_p\}$ for model identification purposes.⁴

2.2 Model Specifications and Estimation

We compare models according to a 2 x 2 x 2 x 2 ‘design’, where the Selection submodel is either Binary or Polychotomous, the Prediction submodel is either Linear-based or Ordinal, ρ is either fixed to zero or estimated, and heterogeneity is allowed for or not. By “linear-based”, we mean a model whose cutoffs are deliberately set apart at constant intervals; because the prediction submodel is identified up to linearity (i.e., location and scale), the actual distance between cutoffs is not important, so we evenly subdivide the unit interval, rendering the results directly comparable to the ordinal cutoff scaling. Note that this is *not* equivalent to the more common OLS-based prediction submodel, which would be an *a priori* misspecification: while it most certainly would treat the dependent variable as linear, it would also presume not only continuity, but boundedness.

Thus, the eight resulting joint models (each of which can include various forms of heterogeneity) are:

<i>Model</i>	<i>Selection</i>	<i>Prediction</i>	<i>Endogeneity</i>	<i>Model</i>	<i>Selection</i>	<i>Prediction</i>	<i>Endogeneity</i>
$M_{BL,0}$	<i>Binary</i>	<i>Linear</i>	$\rho = 0$	$M_{PL,0}$	<i>Poly</i>	<i>Linear</i>	$\rho = 0$
$M_{BL,\rho}$	<i>Binary</i>	<i>Linear</i>	$\rho \neq 0$	$M_{PL,\rho}$	<i>Poly</i>	<i>Linear</i>	$\rho \neq 0$
$M_{BO,0}$	<i>Binary</i>	<i>Ordinal</i>	$\rho = 0$	$M_{PO,0}$	<i>Poly</i>	<i>Ordinal</i>	$\rho = 0$
$M_{BO,\rho}$	<i>Binary</i>	<i>Ordinal</i>	$\rho \neq 0$	$M_{PO,\rho}$	<i>Poly</i>	<i>Ordinal</i>	$\rho \neq 0$

The non-endogenous ($\rho = 0$) models presume that the selection and prediction submodels are independent, and can therefore be estimated separately. They serve as benchmarks, and we will not discuss them other than in this capacity. For the polychotomous selection models, because all comparisons are between-item, there is no estimated constant; for the binary probit selection models, a constant is required. Note that the ordinal models nest the linear ones; the endogenous models nest those where $\rho = 0$; but the polychotomous and binary models are not comparable in this manner.⁵ Finally, any of the models can be estimated

⁴We are indebted to Terry Elrod for this derivation, as well as for the WINBUGS code which exploits it for efficient estimation purposes. Extensive simulation studies – including several using the same number of categories and regressors as in the forthcoming experiment, but varying regressor intercorrelation and degree of error correlation – verified estimation accuracy. Simulation results for all models estimated here are available from the authors.

⁵Polychotomous models will appear to fit dramatically better than binary ones, because the sample space under binary

with and without heterogeneity. For this, we use a hierarchical specification, with a normal random effect for each estimated parameter; for brevity, we include results only for the best-performing models, $M_{BO,\rho}$ and $M_{PO,\rho}$.⁶

We take the somewhat strong position that model $M_{PO,\rho}$ is ‘correct’, in the sense that whatever misspecifications it may be guilty of are present in all the other models as well, so any deviations are due to parameter restrictions (e.g., linear cutoffs, $\rho = 0$) or overt misspecifications (binary probit selection). Note that, for the non-endogenous models, the selection and prediction submodels can be estimated independently, and so corresponding coefficients agree across models.⁷ We have successfully estimated subsets of the 8 models using numerous packages, specifically TSP, LIMDEP, STATA and MATLAB; the polychotomous, endogenous models – $M_{PO,\rho}$ and its linear variant $M_{PL,\rho}$ – were estimated classically in TSP and from a Bayesian perspective in WINBUGS.⁸ We report posterior means for all parameters, noting that standard MLE estimates correspond to a posterior mode under a flat prior, while we used highly diffuse, but weakly informative, conjugate priors for our MCMC runs.

2.3 Model Comparison

Nested non-heterogeneous models can be compared classically through a likelihood-ratio (or Wald) test, and we rely on such tests when they are applicable. However, the heterogeneous forms of the models are not amenable to classical tests, and a variety of methods have appeared in the Bayesian literature. While Bayesian model-testing is straightforward in theory, practice is another matter, with a good deal of debate over numerical recipes for integrated likelihoods, as well as appropriate comparisons for hierarchical models in which, to quote Spiegelhalter et al. (2002), “the number of parameters is not clearly defined”.

Although Damien and Walker (2002) give lie to the “myth” that the most complex of a set of candidate

selection consists of all 32 possible five-tuples with binary entries, while that for polychotomous selection further conditions on there being but a single 1 among them. Thus, they have different observables, and cannot be compared using likelihoods or measures relying on them, such as DIC.

⁶Recent work (e.g., Andrews, Ansari, and Currim 2002; Wedel et al. 1999) comparing finite mixture, hierarchical Bayes and other heterogeneity models indicates that, substantively, they offer roughly equal performance in a variety of empirical settings. In our applications, latent classes led neither to improvements nor alternate conclusions.

⁷Specifically, coefficients agree: for Selection in $\{M_{PO,0}, M_{PL,0}\}$ and $\{M_{BO,0}, M_{BL,0}\}$; and for Prediction in $\{M_{PO,0}, M_{BO,0}\}$ and $\{M_{PL,0}, M_{BL,0}\}$. Thus, the tests for ordinality vs. linearity ($M_{PO,0}$ vs. $M_{PL,0}$, $M_{BO,0}$ vs. $M_{BL,0}$) are identical.

⁸We achieved four significant digit agreement on restricted joint models estimable in at least two of TSP, LIMDEP, WINBUGS and MATLAB; for the last, models were estimated using constrained optimization using both quadrature and simulated likelihood. We further implemented the model using quadrature alone in the spreadsheet program EXCEL which, though hardly ideal, demonstrates the potential for wide applicability. We consider this agreement encouraging evidence of both convergence and accuracy of coefficient estimates for the endogenous models $M_{PO,\rho}$ and $M_{PL,\rho}$. TSP offers an attractive platform for quadrature-based model estimation, as it computes gradients and Hessians analytically from the likelihood function. WINBUGS is a powerful platform for applying MCMC methods, monitoring numerous quantities of interest to the empirical researcher; all hierarchical Bayes models here were estimated in WINBUGS. We wish to thank Clint Cummins of TSP International (<http://www.tspintl.com>) for his help in efficiently programming the likelihood optimizations, and Terry Elrod for lending his expertise in WINBUGS. All code is freely available at [http://\(removed for review\)](http://(removed for review)).

models must fare best, it is very likely to without properly accounting for the number of free parameters in each. Spiegelhalter et al. (1998, 2002), using decision-theoretic arguments, derive the *Deviance Information Criterion* (DIC), which allows comparison across (Bayesian) models of arbitrary complexity. It is readily calculated from MCMC output, and we use it for model comparison when various forms of heterogeneity are accounted for through hierarchical specifications. Because ρ lies in $(-1, 1)$, one cannot assume asymptotic normality for estimates. It is therefore customary (e.g., STATA or LIMDEP) when using selectivity models to report means and standard errors for $\text{atanh}(\rho)$, which takes values on the whole line, along with a corresponding point estimate for ρ . This is expedient when, as in our second data set, we wish to impose a normal hierarchical specification for endogeneity, readily accomplished for $\text{atanh}(\rho)$, but inappropriate for ρ itself.

In comparing models for a particular data set, one point can't be stressed too strongly: the Prediction submodel does not itself change.⁹ Rather, our main concern is how the (often presumed irrelevant) Selection model affects deductions *based on* the Prediction model. Specifically, as we look across models, we will ask to what degree conclusions about effects strengths (i.e., coefficients) for the Prediction model are affected by the specification for the Selection model, in terms of polytomy, heterogeneity and, above all, endogeneity.

3 Empirical Applications

We illustrate the model, and the importance of accounting for choice endogeneity, by applying it to two datasets. The first was collected specifically for this purpose, while the second derives from Ratner, Kahn and Kahneman (1999). Our goal is not to engage in a substantive re-examination of these data, but to use them as a basis for comparison across a set of candidate models. Our descriptions of the data are deliberately concise. We note that the dependent variable differs in the two studies – ordinal in the first case, continuous (linear or interval-scaled) in the second – but we shall not otherwise distinguish among data types for observables except where necessary.

3.1 Application I: Effects of Choice Set Composition on Evaluation

We examine data on choices among candy bars from a pilot study conducted by Brown (2003). The experiment was described as a “candy preferences study” in which the researchers were interested in what types of candy were preferred by the university’s students. Seventy-eight subjects were asked to rate each

⁹This is true, subject to the minor exception of whether, for the ordered categorical observable, cutoffs are estimated or evenly spaced.

of ten popular candies on a 1 - 7 scale with a hedonically neutral mid-point (Kahneman, Wakker and Sarin 1997; “dislike very much” - “like very much”), yielding a measure of subjects’ prior preferences for each item; rankings were collected as well. During an unrelated 30-minute filler task, each subject’s pre-test response was used to create a customized choice set of actual candy bars, based on his or her ratings. It was important that not all subjects choose a favorite, so a subject’s two top-ranked items were never included in his/her set (thus the item ranked third in the pre-test was always a “favored item”). Each subject was then presented with either a small (two-item) or a larger (five-item) set, asked to choose one candy bar, consume it, and rate it on the same (ordinal) 1 - 7 scale. A variety of checks (regarding the cover story, error correlation, and the accuracy of the ‘prior’ ratings) indicated that the experimental manipulations operated as intended.

The (ordinal) dependent variable, POST, is the subject’s posterior rating, and it is observed only for the single item chosen. Regressors for the selection model (for which brand is chosen) are HIGHRANK (whether the brand had the highest *a priori* rank) and DISTANCE (distance of each brand to the highest-ranked, in number of rating points); the single regressor for the prediction model (for the single brand chosen) is DISTLARG (distance from the chosen brand to the lowest-rated, which is greater in the larger choice sets, by design). A number of other covariates are available as well, but substantively altered the forthcoming results only to the extent that they were highly collinear with some subset of the predictors above; as such we present this regressor specification alone.¹⁰ TABLE 1 lists estimates for the eight non-heterogeneous models, as well as for heterogeneous versions of the ‘best’ models, $M_{BO,\rho,Heterog}$ and $M_{PO,\rho,Heterog}$.

[TABLE 1 about here]

3.2 Comparing Implications Across Models

A number of intriguing differences deriving from the various selection and prediction submodel types are apparent from TABLE 1. It is possible to test whether cutoffs are linearly (i.e., equally) spaced, akin to the interval-like scaling presumed by the OLS portion of the standard Heckman formulation (1)-(3). As can be seen from TABLE 1, such linearity can be rejected in all cases.

Substantively, however, when focusing only on the effect estimate for the Prediction model, $\hat{\beta}_{DISTLARG}$, the pattern across the polychotomous cases ($-0.159, -0.454, -0.080, -0.385$) and that across the binary ones ($-0.159, -0.516, -0.080, -0.489$) appear quite similar. More troubling are systematic upward biases

¹⁰Importantly, endogeneity was present in all such models. Although the extent, as measured by ρ , did vary across the set of possible models, it did not do so significantly, and was always statistically significant taken on its own. This, in fact, argues against an ‘omitted variables’ type explanation for our results.

in $\hat{\beta}_{\text{DISTLARG}}$ when the selection method is (incorrectly) presumed binary: estimates are inflated substantially (Linear: -0.516 vs. -0.454 ; Ordinal: -0.489 vs. -0.385), and we report informally that similar biases were apparent in our simulation studies.

Of all potential misspecifications, however, ignoring endogeneity entails the greatest possibility of drawing incorrect conclusions. In all cases, the $\rho = 0$ hypothesis can be strongly rejected (based on the Wald test). In comparisons between the $\rho = 0$ selection submodel (i.e., no selection effects at all) and its endogenous variant, these differences are pronounced: for the polychotomous ordinal models, allowing for endogeneity ($M_{PO,\rho}$) yields $\hat{\beta}_{\text{DISTLARG}} = -0.489$, a significant effect; presuming there is no endogeneity ($M_{PO,0}$) yields $\hat{\beta}_{\text{DISTLARG}} = -.080$, a negligible one. For these data, *failing to account for endogeneity leads to the incorrect conclusion that the prediction model was essentially useless*. Although these data indicate a rather high degree of endogeneity, $\rho \approx .75$, even with a small sample size, evidence of its presence is exceptionally strong.

Some prior research (e.g., Chintagunta 2001) has suggested that failing to account for heterogeneity can lead one to misestimate – usually, to overestimate – the degree of endogeneity present; but this isn’t so for these data. A (normal) hierarchical specification for coefficient heterogeneity suggests that there is a good deal of variation in $\hat{\beta}_{\text{DISTLARG}}$, with a mean of -0.474 and a standard deviation of 1.082 . However, the size and significance of ρ are hardly affected: for the best model ($M_{PO,\rho}$) coefficient heterogeneity changes ρ trivially, from 0.791 to 0.766 .

One might question which is more ‘valuable’ here, heterogeneity or endogeneity. While any answer is speculative in such a small data set, we can compare DIC values for the polychotomous, ordinal models. With neither endogeneity nor heterogeneity, $\text{DIC} = 339.8$; with only heterogeneity, $\text{DIC} = 297.0$; with only endogeneity, $\text{DIC} = 277.8$; and with both, $\text{DIC} = 261.5$. It would thus appear that, of the two, endogeneity reduces DIC considerably more than heterogeneity, although we have not systematically confirmed this by trying many alternate forms of heterogeneity correction.

Overall, analysis of these data provides evidence that each of the constructs involved in the proposed model – polychotomous selection, ordinal prediction and, above all, choice endogeneity – adds substantially to both model fit and interpretation of effects strengths. Moreover, heterogeneity isn’t the root cause of these findings; it neither diminishes nor amplifies them substantially when it’s accounted for. We next take up the issue of whether ρ itself differs across respondents, by examining data in which there are multiple observations for each.

3.3 Application II: Preference for Varied Experience

A number of studies (c.f., Simonson 1990) suggest that consumers can behave as if varied sequences – of products or other experiences – were intrinsically superior to repetitive ones. And, so the logic goes, they will consume items they are less fond of to achieve that ‘variedness’. Contrary to the hypothesis that one’s tendency to choose less-preferred items stems from satiation on frequently-consumed favorite items, Ratner, Kahn, and Kahneman (1999; henceforth RKK) demonstrated that more varied sequences of popular songs resulted in *diminished* enjoyment during consumption (even though participants “did not become satiated with top-ranked songs”). RKK’s analysis made use of straightforward statistical methods and, because the focus of their paper did not include choice endogeneity, no Heckman- or Tobit-type corrections were considered.

RKK (Experiment 5) studied the effects of variety and frequency of consumption on ratings of popular songs.¹¹ Each of 59 subjects provided a sequence of 20 real-time ratings, on a 100-point scale, for 12 popular songs that were presented to them through a computer program. Subjects were first presented with either a set of 3 items (ranked 1,3,6 or 2,4,7 according to subjects’ prior ratings) or a set of 6 (those ranked 2,4,7,10,11,12 or 1,3,6,8,9,10), chose, listened to and rated the song of their choice, also on a 100-point scale; this was done 10 times. For an additional 10 occasions, subjects were presented with the other-sized choice set. Because these data involve two phases – choice (of which song one listens to) from sets of varied sizes, and ratings (of satisfaction or liking of the chosen song) – it is perfectly suited to the models developed here, and the potential for choice endogeneity.

Our re-analysis is narrowly focused, on whether *the act of having chosen* (Selection) affects the degree to which one is satisfied with one’s choice (Prediction). To this end, we predict posterior ratings (RATINGS) using prior ones (PRIOR) and the size of the choice set (VARIETY: either 3 or 6 songs), while ‘endogenously correcting’ for the choice process by accounting for whether a favorite was chosen (FAVORITE) and, again, prior ratings (PRIOR).¹² The only change in the model involved the form of the observed variable, RATINGS, which lies on a 100-point scale, and so may be taken as continuous, as opposed to the ordinal scaling of the previous example. Although it is possible to appeal to a parsimonious specifications for ordinal variables

¹¹We thank the authors for allowing us to re-analyze their experimental data.

¹²Note that it is acceptable to include the same covariates in the Selection and Prediction submodels, particularly so when, as here, theory suggests doing so. In particular, omitting PRIOR in the selection model might literally *create* a large value of ρ . Note, however, that it is impossible to include VARIETY in the selection model, as it takes constant values within each choice set, and so can only drive evaluations between them, not within them. Many alternative forms are possible for either model; the one presented here is that among those proposed by theory which offered the most parsimonious pattern of significant effects. Endogeneity, as measured by ρ , was significant in all such models.

– such as Rost’s (1985) rank-order binomial set-up – we find no empirical differences between this and the more natural continuous specification, and so make use of the latter.¹³ Model estimates are given in TABLE 2.

[TABLE 2 about here]

We do not engage in a substantive re-interpretation of the very rich RKK data, which also contains a temporal component which we do not model here, not to mention the numerous companion studies in that paper. We can, however, consider model implications strictly from the vantage point of endogeneity, which appears to be overwhelmingly present in these data, regardless of which model is used. Whether this is truly the effect of classic endogeneity – caused by allowing subjects to choose their own songs to rate – or a more subtle effect of some omitted variable, is open to question. However, endogeneity was strongly indicated in a variety of other regressor specifications for both the prediction and selection submodels, arguing against the latter possibility.

We include two examples of (heterogeneous) binary selection models for comparison purposes; because they are overtly misspecified, fit worse and entail incorrect signs for some of the Prediction effects, we won’t refer to them again. We discuss five forms of the polychotomous model; in order of their appearance in TABLE 2: (1) neither endogeneity nor heterogeneity; (2) β -heterogeneity only; (3) endogeneity only; (4) endogeneity and β -heterogeneity; (5) endogeneity, β -heterogeneity and ρ -heterogeneity. In this way, one can ‘decompose’ the influences of the various modeling constructs systematically.

Substantively, the selection model is of lesser interest, and effect strengths across the various polychotomous models (listed RKK1-5 in TABLE 2) are very close; they in fact are not statistically distinguishable. A very different story emerges across the Prediction models: while the effects of PRIOR ratings are always quite strong (and of roughly equal strength), the effects of VARIETY, the main construct under study, are *vastly* different. When there is no endogeneity ($\rho = 0$, RKK1), the effects of VARIETY are very weakly negative, $\hat{\beta}_{\text{VARIETY}} = -0.063$ (ns). That is, if one analyzed only the RATINGS (i.e., Prediction) data, choice set size could be confidently claimed as insubstantial. However, when endogeneity is accounted for ($\rho \neq 0$, RKK3), we see that VARIETY has a very strongly positive effect, $\hat{\beta}_{\text{VARIETY}} = 2.605$. *A posteriori*, in comparing a

¹³It is not difficult to accommodate different variable types for the Prediction submodel. When observations are continuous, a subject’s likelihood contribution is:

$$\frac{1}{\bar{\rho}} \int_{\theta \in \mathbb{R}} \phi(\theta) \left(\prod_{i>1} \Phi[(U_{s,[1:k]} - U_{s,[i:k]}) + \theta] \right) \phi \left[\frac{(y - U_{p,[1:k]}) - \sigma \rho \theta}{\sigma \bar{\rho}} \right] d\theta.$$

choice set of size 3 to one of size 6 (as RKK did), *chosen items are rated nearly 8 points higher when chosen from the larger set*. An important effect is therefore entirely missed when the Ratings data are analyzed in the absence of an associated model for choice.¹⁴

3.3.1 The Effects of Heterogeneity in β and in ρ

One can legitimately ask to what extent this “important effect” may be an artifact of heterogeneity; perhaps some people react strongly, while others don’t, or perhaps only some exhibit endogeneity at all. Thus, we consider the analogs of the two previous models (RKK1, 3) when β -heterogeneity is introduced (RKK2, 4) and when ρ -heterogeneity is further introduced (RKK5). On its own, β -heterogeneity doesn’t reverse the pattern of results: the mean of $\hat{\beta}_{\text{VARIETY}}$ is 0.026 with no endogeneity (RKK2), but 1.744 when it is accounted for (RKK4). In both cases, however, there is a great deal of heterogeneity, as the standard deviation – which estimates the variation across respondents – is about 2.7 in each case. It would therefore appear that response to VARIETY is itself quite varied. Note that none of these quantities is very sharply estimated; while there is a great deal of data (59 subjects, 20 choices each), there were relatively few subjects.

The presence of multiple choices per subject allows us to assess whether the degree of endogeneity itself varies across subjects. As discussed previously, we accomplish this by placing a normal hierarchical specification on $\text{atanh}(\rho)$, as well as on β . The results are intriguing: although the mean degree of endogeneity (ρ) doesn’t change (and is, in fact, very strong), allowing it to vary across subjects has an obvious impact on VARIETY (though not so on PRIOR, nor on the Selection model itself). Specifically, when both β - and ρ -heterogeneity are allowed in the endogenous model (RKK5), the mean of the effect for VARIETY drops further, to 1.085, a great reduction from β -heterogeneity alone (RKK4; $\hat{\beta}_{\text{VARIETY}} = 1.744$) and the non-heterogeneous model (RKK2; $\hat{\beta}_{\text{VARIETY}} = 2.605$). We can conclude that lack of appropriate heterogeneity correction, in both β and in ρ , appears to inflate mean effects estimates.

One is left with the question of which model represents the data best and, on this, both LL (using the harmonic mean estimator) and DIC (which is preferred for model selection) are unequivocal: the value of the latter, for RKK1-5, are {12727, 12059, 9792, 8870, 8268}. The most complex model thus more than compensates for its additional complexity, and both forms of heterogeneity are important. We note in closing that the non-endogenous models fit exceedingly poorly, as indicated by DIC and the large value of ρ in all endogenous versions of the model.

¹⁴While space precludes the presentation of numerous other alternative specifications, we can report that this finding is robust to the inclusion of additional predictors in either model, as well as standard transforms.

4 Conclusions and Potential Extensions

Selectivity models developed by Heckman, Tobin, and others have clarified the effects of failing to account for endogeneity. However, these models are restricted to the form most typically faced in the economics literature – binary selection coupled with continuous, interval-scaled prediction. Our intent in this paper has been to demonstrate that similar selectivity mechanisms are intrinsic to a broad class of decision problems common in consumer and psychological research, and to show how researchers can account for them in a general setting.

Although we have not reported on them in this paper, we have successfully extended the model to allow for different types of selection and prediction types, including ‘pick-any’ and ranked selection (i.e., the field is narrowed not to just one, but to several, options, which may be ranked) and discrete choice prediction (i.e., we observe only what was finally chosen, but not any rating or evaluation of it). A common example of such extensions is the process of purchasing a car, which typically involves several distinct phases – information-gathering, visiting dealers, test driving – before a choice is made. Researchers ignoring the individual-specific selection phase(s) preceding eventual choice may be led astray in gauging just what drives the market. For example, price may determine which cars are eliminated early on, and may thus appear relatively unimportant if only later stages of the purchase process are analyzed. Applying an appropriate member of this class of models would allow one to disentangle the effects of (perhaps multiple phases of) selection, as well as a discrete choice prediction, while accounting for multiple sources of heterogeneity.

We can envision several fruitful extensions of the basic methodology. For example, ‘selection’ in our model requires full knowledge of available options and item covariates, which are rarely available in field data, and require care and foresight in experimental settings. We have also not explored, other than by trial and error, how one chooses the best regressor set for models with an endogenous selection phase, and whether analogs to stepwise or ridge regression might be useful. Although our hierarchical formulation of heterogeneity is fairly general, we cannot claim that it works well beyond the two data sets studied here. Nevertheless, we do describe a class of endogenous extensions to standard models broadly applicable to studies of judgment and decision-making, and hope that their ease of use will encourage experimental researchers to apply them.

5 Appendix: Model Identification and Likelihoods

We wish to show that there is a ‘canonical’ scaling for the multinomial probit (selection) model which allows conditioning on a single (independent) standard normal variate to calculate model likelihoods.¹⁵ Let us consider the standard set-up for an MNP error process (e.g., McCulloch, Polson and Rossi 2000). The probability that Item 1 is in fact chosen would then be:

$$P \left[U_{s,[1:k]} + \epsilon_{s,[1:k]} > \{U_{s,[i:k]} + \epsilon_{s,[i:k]}\}_{i>1} \mid \{\epsilon_s\} \sim N_k [0, \Sigma_\epsilon] \right], \quad (6)$$

with Σ_ϵ positive definite. It is well known that estimating the elements of Σ_ϵ entails (linear) scale and location indeterminacies, in that a constant can be added to each of the $U_{s,[i:k]}$, and the entire system can be divided by any positive constant. We wish to re-write the error process $\{\epsilon_s\}$ in terms of an independent, standard normal draw for $\epsilon_{s,[1:k]}$, and a non-degenerate MVN error process for $\{\epsilon_{s,[i:k]}\}_{i>1}$. Because Σ_ϵ is positive definite, all eigenvalues are positive, so we consider the smallest, and call it λ . By definition, this solves the characteristic equation $\det(\Sigma_\epsilon - \lambda I) = 0$. It further shows that $(\lambda^{-1}\Sigma_\epsilon - I)$ is singular and, because λ is the *smallest* eigenvalue, the eigenvalues of $(\lambda^{-1}\Sigma_\epsilon - I)$ must be positive, except for a single zero. Therefore, we can rescale *all* quantities in (6) by $\lambda^{-1/2}$, so that we have independent error processes, $\{\nu_s\} \sim N_k [0, I]$ and $\{\xi_s\} \sim N_{k-1} [0, \Sigma_\xi]$, and Σ_ξ is the sub-matrix of $(\lambda^{-1}\Sigma_\epsilon - I)$ formed by deleting the first row and column:

$$P \left[\lambda^{-1/2} U_{s,[1:k]} + \nu_{s,[1:k]} > \left\{ \lambda^{-1/2} U_{s,[i:k]} + \nu_{s,[i:k]} + \xi_{s,[i:k]} \right\}_{i>1} \right]. \quad (7)$$

This provides the independent standard normal variate $\nu_{s,[1:k]}$ as the sole error component for the focal item in (7). It can then be unambiguously correlated with the error component in the Prediction model, $\epsilon_{p,[1:k]}$. Further, because the error process $\{\xi_s\}$ is MVN with (positive definite) covariance matrix Σ_ξ , and is independent of $\{\nu_s\}$, MCMC can proceed with ordinary Wishart draws. Lastly, for the special case when errors are independent, we can skip MCMC steps for $\{\xi_s\}$ entirely, and (7) reverts to the form used in the likelihood, (5).

¹⁵We are grateful to Terry Elrod for proposing this derivation and general approach.

References

- [1] Agresti, A. (1990), *Categorical Data Analysis*, John Wiley & Sons, New York.
- [2] Amemiya, Takeshi (1985), *Advanced Econometrics*, Harvard University Press, Cambridge, MA.
- [3] Andrews, Rick L., Asim Ansari, and Imran S. Currim (2002), “Hierarchical Bayes Versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery,” *Journal of Marketing Research*, 39 (February), 87-98.
- [4] Berry, Steven, James Levinsohn and Ariel Pakes (1995), “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841-90.
- [5] Besanko, D., S. Gupta and D. Jain (1998), “Logit Demand Estimation Under Competitive Pricing Behavior: An Equilibrium Framework,” *Management Science*, 44(11), 1533–1547.
- [6] Blundell, R. and J.L. Powell (2003), “Endogeneity in Nonparametric and Semiparametric Regression Models,” in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II* (Cambridge University Press).
- [7] Boatwright, Peter, Sanjay Dhar and Peter Rossi (2003), “The Role of Retail Competition and Account Retail Strategy as Drivers of Promotional Sensitivity”, *Journal of Business*, forthcoming.
- [8] Braun, Kathryn A. (1999), “Postexperience Advertising Effects on Consumer Memory,” *Journal of Consumer Research*, 25 (March): 219-344.
- [9] Bronnenberg, Bart J. and Vijay Mahajan (2001), “Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables,” *Marketing Science*, 20(3), Summer, 284-299.
- [10] Brown, Christina (2003), “If You’ve Got A Lemon, Make Lemonade: A Preference-Formation Explanation For Comparison Effects In Judgment and Choice,” working paper, University of Michigan Business School.
- [11] Chintagunta, P. K. (2001), “Endogeneity and Heterogeneity In A Probit Demand Model: Estimation Using Aggregate Data,” *Marketing Science*, 20(4), Fall, 442-456.
- [12] Chintagunta, P. K., V. Kadiyali, V. and N. Vilcassim (1999), “Endogeneity and Simultaneity in Competitive Price and Advertising: A Logit Demand Analysis,” working paper, University of Chicago.
- [13] Dawes, R. M. (1979), “The Robust Beauty of Improper Linear Models in Decision Making,” *American Psychologist*, 34, 571-582.
- [14] Degeratu, Alexandru M., Arvind Rangaswamy and Jianan Wu (2001), “Consumer Choice Behavior in Online and Traditional Supermarkets: The Effects of Brand Name, Price, and Other Search Attributes,” *International Journal of Research in Marketing*, 17(1), 55-78.
- [15] Draganska, M. and D. Jain (2003), “Structural Models of Competitive Market Behavior: An Estimation Approach Using Disaggregate Data,” working paper, Northwestern University.
- [16] Dubin, Jeffrey A. and Daniel L. McFadden (1984), “An Econometric Analysis of Residential Electric Appliance Holdings and Consumption,” *Econometrica*, 52(2), 345-62.
- [17] Erdem, Tülin, Susumu Imai and Michael P. Keane (2003), “A Model of Consumer Brand and Quantity Choice Dynamics under Price Uncertainty,” *Quantitative Marketing and Economics*, 1(1), 5-64.

- [18] Feinberg, Fred M. and Joel Huber (1996), "A Theory of Cutoff Formation Under Imperfect Information," *Management Science*, 42(1), 65-84.
- [19] Fitzsimons, Gavan (2000), "Consumer Response to Stockouts," *Journal of Consumer Research*, 27 (September), 249-266.
- [20] Häubl, G. and V. Trifts (2000), "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Marketing Science*, 19(1), 4-21.
- [21] Hay, J. W. (1980), "Occupational Choice and Occupational Earnings: Selectivity Bias in a Simultaneous Logit-OLS Model," Unpublished Ph.D. Dissertation, Yale University.
- [22] Heckman, James J. (1976), "Common Structure of Statistical-Models of Truncation, Sample Selection And Limited Dependent Variables And a Simple Estimator For Such Models," *Annals of Economic And Social Measurement*, 5(4), 475-492.
- [23] Heckman, James J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153-161.
- [24] Honore, B. E. and L. Hu (2002), "Estimation of Cross Sectional and Panel Data Censored Regression Models with Endogeneity," working paper, Princeton University.
- [25] Huang, Chun-Yao, Bruce G. S. Hardie and William Putsis (2001), "The Impact of Sample Selection on Measures of Price Elasticity from Store-Level Market Response Models," working paper, London Business School.
- [26] Iyengar, S. and M. Lepper (2000), "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?," *Journal of Personality and Social Psychology*, 79(6), 995-1006.
- [27] Kahneman, D., P.P. Wakker, and R. Sarin (1997), "Back to Bentham? Explorations of Experienced Utility," *Quarterly Journal of Economics*, 112, May, 375-405.
- [28] Lee, L. F. (1983), "Generalized Econometric Models With Selectivity," *Econometrica*, 51.
- [29] Levin, Irwin P., J. D. Jasp and Wendy S. Forbes (1998), "Choosing Versus Rejecting Options at Different Stages of Decision Making," *Journal of Behavioral Decision Making*, 11: 193-210.
- [30] Levin, Irwin P., Caryn M. Prosansky, Daniel Heller, and Brad M. Brunicck (1991), "Prescreening of Choice Options in 'Positive' and 'Negative' Decision-Making Tasks," *Journal of Behavioral Decision Making*, 14: 279-293.
- [31] Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.
- [32] McCulloch, R., N. Polson and P. Rossi (2000), "Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters," *Journal of Econometrics*, 99, 173-193.
- [33] Nawata, Kazumitsu (1994), "Estimation of Sample Selection Bias Models by the Maximum Likelihood Estimator and Heckman's Two-Step Estimator," *Economics Letters*, 45, 33-40.
- [34] Nelson, F. D. (1984), "Efficiency of the 2-Step Estimator For Models With Endogenous Sample Selection," *Journal of Econometrics*, 24 (1-2), 181-196.
- [35] Puhani, Patrick A. (2000), "The Heckman Correction For Sample Selection And Its Critique," *Journal of Economic Surveys*, 14(1), 53-68.

- [36] Ratner, Rebecca K., Barbara E. Kahn and Daniel Kahneman (1999), "Choosing Less-Preferred Experiences for the Sake of Variety," *Journal of Consumer Research*, 26(1), 1-15.
- [37] Reiss, Peter C. and Frank Wolak (2002), "Structural Econometric Modeling," *Handbook of Econometrics*, Volume 5 (forthcoming).
- [38] Rost, J. (1985), "A Latent Class Model for Rating Data," *Psychometrika*, 50, 37-49.
- [39] Shiv, Baba and Joel Huber (2000), "The Impact of Anticipating Satisfaction on Consumer Choice," *Journal of Consumer Research*, 27 (September): 202-216.
- [40] Simonson, Itamar (1990), "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior," *Journal of Marketing Research*, 27 (May), 150-162.
- [41] Spiegelhalter, D.J., N.G. Best and B.P. Carlin (1998), "Bayesian Deviance, the Effective Number of Parameters, and the Comparison of Arbitrarily Complex Models," Technical Report, MRC Biostatistics Unit, Cambridge, UK.
- [42] Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A. van der Linde (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society, Series B*, 64, 583-639.
- [43] Tobin, James (1958), "Estimation of Relationships Among Limited Dependent Variables," *Econometrica*, 26, 24-36.
- [44] Tversky, Amos (1972), "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79, 281-299.
- [45] Villas-Boas, M. and R. Winer (1999), "Endogeneity and Brand Choice Models," *Management Science*, 45(10), 1324-1338.
- [46] Wedel, Michel, Wagner A. Kamakura, Albert C. Bemmaor, J. Chiang, Terry Elrod, R. Johnson, Peter J. Lenk, Scott A. Neslin, and C. S. Poulsen (1999), "Discrete and Continuous Representation of Heterogeneity," *Marketing Letters*, 10 (3), 217-30.
- [47] Winship, Christopher and Robert D. Mare (1992), "Models For Sample Selection Bias," *Annual Review of Sociology*, 18, 327-350.
- [48] Zhang, Shi and Gavan Fitzsimons (1999), "Choice Process Satisfaction: The Influence of Attribute Alignability and Option Limitation," *Organizational Behavior and Human Decision Processes*, 77 (3), 192-214.

Table 1

Candy Data Model Comparisons

Posterior Means for Selection, Prediction, Endogeneity (ρ), Heterogeneity in β

Selection Model	Binary				Polychotomous			
	Linear	Ordinal	Ordinal	Ordinal	Linear	Ordinal	Ordinal	Ordinal
Prediction Model	---	Yes	---	Yes	---	Yes	---	Yes
Endogeneity (ρ)	---	---	---	Yes	---	---	---	Yes
Heterogeneity in β	---	---	---	Yes	---	---	---	Yes
Model	$M_{BL,0}$	$M_{BL,p}$	$M_{BO,0}$	$M_{BO,p}$	$M_{PL,0}$	$M_{PL,p}$	$M_{PO,0}$	$M_{PO,p}$
Selection Model	mean		std		mean		std	
HiRank	1.025	0.912	1.025	0.943	0.581	0.408	0.581	0.443
stderr	0.212	0.172	0.212	0.178	0.243	0.226	0.243	0.228
Distance	-0.240	-0.322	-0.240	-0.312	-0.305	-0.472	-0.305	-0.436
stderr	0.078	0.064	0.078	0.066	0.107	0.117	0.107	0.112
C	-0.700	-0.199	-0.700	-0.212				
stderr	0.155	0.123	0.155	0.126				
Prediction Model	mean		std		mean		std	
DistLarg	-0.159	-0.454	-0.080	-0.385	-0.159	-0.516	-0.080	-0.489
stderr	0.249	0.234	0.303	0.285	0.249	0.225	0.303	0.288
C	0.407	0.294	0.518	0.371	0.407	0.292	0.518	0.369
stderr	0.026	0.034	0.054	0.068	0.026	0.032	0.054	0.064
σ	0.209	0.236	0.246	0.273	0.209	0.220	0.246	0.261
stderr	0.018	0.024	0.028	0.035	0.018	0.020	0.028	0.031
Cutoffs	0		0		0		0	
μ_1	0	0	0	0	0	0	0	0
μ_2	0.2	0.2	0.212	0.198	0.2	0.2	0.212	0.201
μ_3	0.4	0.4	0.553	0.518	0.4	0.4	0.553	0.530
μ_4	0.6	0.6	0.800	0.773	0.6	0.6	0.800	0.781
μ_5	0.8	0.8	0.910	0.894	0.8	0.8	0.910	0.899
μ_6	1	1	1	1	1	1	1	1
Endogeneity	0.772		0.706		0.835		0.791	
ρ	1.076		0.924		1.363		1.205	
atanh(ρ)	0.260		0.255		0.485		0.457	
stderr	0.490		0.490		0.487		0.487	
Model Comparison and Tests	-243.24		-233.17		-235.13		-227.64	
Classical LL	-186.29		-176.78		-178.18		-170.65	
-2 Δ (LL) for Ordinality, 4df	16.22		11.06		16.22		12.26	
p-value for Ordinality	2.7E-03		2.6E-02		2.7E-03		1.6E-02	
-2 Δ (LL) for Endogeneity, 1df	20.14		14.98		19.02		15.06	
p-value for Endogeneity	7.2E-06		1.1E-04		1.3E-05		1.0E-04	

