A Discrete Choice Model for Partially Ordered Alternatives.*

Eleni Aristodemou†
University of Cyprus

Adam M. Rosen‡
Duke University, CeMMAP and IFS

January 17, 2022

Abstract

In this paper we analyze a discrete choice model for partially ordered alternatives. The alternatives are differentiated along two dimensions, the first an unordered "horizontal" dimension, and the second an ordered "vertical" dimension. The model can be used in circumstances in which individuals choose amongst products of different brands, wherein each brand offers an ordered choice menu, for example by offering products of varying quality. The unordered-ordered nature of the discrete choice problem is used to characterize the identified set of model parameters. Following an initial nonparametric analysis that relies on shape restrictions inherent in the ordered dimension of the problem, we then provide a specialized analysis for parametric specifications that

*We are grateful to the co-editor and three anonymous referees for suggestions that led to substantial improvements in this paper. We thank Tim Christensen, Allan Collard-Wexler, Francesca Molinari, Lars Nesheim and several conference and seminar audiences for helpful discussions and comments. Xinyue Bei, Khuong (Lucas) Do, and Muyang Ren provided excellent research assistance. We gratefully acknowledge financial support from the UK Economic and Social Research Council through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (CeMMAP) and through the funding of the "Programme Evaluation for Policy Analysis" node of the UK National Centre for Research Methods, as well as from the European Research Council (ERC) under grants ERC-2009-StG-240910-ROMETA and ERC-2009-AdG, grant agreement number 249529. Eleni Aristodemou gratefully acknowledges financial support from ESRC and UCL. Data supplied by Kantar UK Ltd. The use of Kantar UK Ltd. data in this work does not imply the endorsement of Kantar UK Ltd. in relation to the interpretation or analysis of the data. All errors and omissions remain the responsibility of the authors.

†Address: Department of Economics, Faculty of Economics and Management, University of Cyprus, P.O.Box 20537, CY1678 Nicosia, Cyprus. Email: aristodemou.d.eleni@ucy.ac.cy.

‡Address: Department of Economics, Duke University, 213 Social Sciences, Box 90097, Durham, NC, United States. Email: adam.rosen@duke.edu.

generalize common ordered choice models. We characterize conditional choice probabilities as a function of model primitives with particular analysis focusing on cases in which unobservable taste for quality of each brand offering is multivariate normally distributed. We provide explicit formulae used for estimation and inference via maximum likelihood, and we consider inference based on Wald and quasi-likelihood ratio statistics, the latter of which can be robust to a possible lack of point identification. An empirical illustration is conducted using data on razor blade purchases in which each brand has product offerings vertically differentiated by quality.
JEL classification: C01, C31, C35.

1 Introduction

In this paper we study a discrete choice model in which alternatives are distinguished by two dimensions. The alternatives are first horizontally differentiated according to one of a number of unordered categories. In the context of a consumer choice problem the alternatives could be products differentiated by brands $b = 1, ..., \bar{b}$. Within each such category, alternatives are vertically differentiated by quality $y = 1, 2, ..., \overline{y}_b$. Individuals are assumed to have ordered preferences over the vertical quality dimension, *within* each horizontally differentiated category, but preferences *across* horizontal categories are unordered.

There are several real-world examples of product markets that feature multiple firms competing to sell vertically differentiated alternatives to consumers. For example, airlines sell seat tickets for routes that are vertically differentiated by travel class. Cable and streaming television packages offered by different providers are often vertically-differentiated, with more expensive offerings providing more viewing options or live channels. Cellular phone providers offer vertically differentiated data and cellular plans. Competing ride-share companies offer regular and premium transportation services. Automobile dealers sell quality differentiated cars, for example luxury versus economy, offered by horizontally differentiated manufacturers. These can all lead to the type of vertically and horizontally differentiated choice menus that our model aims to capture.

As initially set out by McFadden (1974), and as is now standard in the discrete choice literature, we assume that each consumer chooses the brand-quality combination that maximizes her latent utility. Our model differs from standard models of discrete choice by explicitly incorporating both the horizontal and vertical dimensions of differentiation. Models that consider choice amongst unordered discrete alternatives, such as those of McFadden

(1974) and Hausman and Wise (1978), allow for horizontal differentiation by brand but do not incorporate vertical differentiation. Models for choice amongst totally ordered alternatives can be used to estimate demand for vertically differentiated products, as in Bresnahan (1987). Nested choice models such as that of Goldberg (1995) allow flexible substitution patterns across nests, but feature unordered (e.g. logit) differentiation within nests.

Our goal here is to combine features of models for ordered and unordered choice in order to incorporate both horizontal and vertical aspects of differentiation. Relative to existing methods, this approach allows the model to reflect the unordered-ordered nature of the choice problem when both kinds of differentiation are present. This may be useful for accurately estimating important features of substitution patterns in such scenarios.

A related line of research, and an important area of potential application, is the modeling of consumer choice in oligopoly markets in which competing firms each offer vertically differentiated products. Some empirical work in this area includes Davies, Waddams, and Wilson (2009) and Song (2015). Davies, Waddams, and Wilson (2009) focus on two-part tariffs and bundling in the British gas and electricity markets, and use linear panel data regression and instrumental variables to investigate whether the market operates in accord with economic theory. Song (2015) develops an explicit model of consumer demand for vertically and horizontally differentiated products, but our model and Song's model are quite distinct and suited for different contexts. Song's (2015) model is a hybrid of those of Berry, Levinsohn, and Pakes (1995) and Berry and Pakes (2007) and is well-suited to settings where products span multiple markets. Moreover, Song (2015) models demand for attributes in characteristics space, and is thus capable of handling a large product space. Our model is instead focused at the consumer level, requiring individual-specific choice data, and is best suited to competition among relatively few brands, or firms, with vertically differentiated product offerings, where demand depends on individual characteristics as well as on product attributes.

In our model, if attention is restricted to any single brand b, the quality of the utility-maximizing option offered by that brand for a given consumer is determined by a standard ordered choice structure. That is, the shape of the latent utility function results in an ordered choice model, e.g. ordered probit or logit, when consumers' choices are restricted to brand b. From a modeling standpoint, this can be used to recover an indirect utility function for each brand b. The solution to the problem of choosing the best brand-quality offering from among all products can then be recovered as the brand that maximizes the indirect utility function, and the quality level that maximizes the corresponding brand-specific utility.

The structure of the problem is thus analogous to that of the mixed discrete-continuous choice model of Dubin and McFadden (1984). However, due to the discrete nature of both dimensions of choice, one cannot use differential arguments and in particular Roy's Identity to characterize the optimal choice of either dimension. Nonetheless, the model is complete in that conditional on any value of exogenous variables, there is a unique solution to the consumer choice problem with probability one. This is because the model is for a single-agent decision problem, rather than a simultaneous move game with strategic interactions and multiple equilibria, as encountered for instance in the simultaneous equations model for ordered actions considered by Aradillas-Lopez and Rosen (2022).

The paper proceeds as follows. In Section 2 we lay out our econometric model for partially ordered response in its most general form, providing shape restrictions on individuals' underlying utility functions that deliver the within-brand ordered choice structure. In Section 3 we characterize conditional choice probabilities (CCPs) and provide identification analysis. The most general formulation of these restrictions offers minimal requirements for this structure but will in general yield set identification, where the identified set of model primitives can be characterized as the set of maximizers of the expected log-likelihood. In Section 4 we consider additional parametric restrictions that preserve the unordered-ordered nature of the underlying choice problem. These lend tractability to the log-likelihood and narrow the size of the identified set, and may result in point identification. Section 5 develops a tractable characterization of CCPs in a parametric partially ordered probit model with two brands, each having two quality tiers, as encountered in our application. This characterization aids in computation, as it can be used to avoid simulation or explicit numerical integration for computing CCPs. Section 6 presents an application to the market for women's razor blades using consumer data from Britain in the early 2000s. Here we describe our empirical implementation and present confidence intervals employing both Wald and QLR tests, the latter having been shown to be robust to partial identification by Chen, Christensen, and Tamer (2018) under certain conditions not formally verified here.[1] Section 7 concludes. Additional material is provided in the Appendices, as referenced in the paper. Proofs of propositions and theorems are provided in Appendix A.

[1]Further results on the distribution of likelihood ratio statistics when point identification fails include those of Liu and Shao (2003) for parametric likelihood models and Chen, Tamer, and Torgovitsky (2011) for semi-parametric models.

2 The Model

Each individual in the population is characterized by observables (Y, B, Z) and an unobservable vector V. It is assumed that each individual chooses either an ordered alternative $Y \in \mathcal{Y}_b \equiv \{1, ..., \bar{y}_b\}$ of some type $B \in \mathcal{B} \equiv \{1, ..., \bar{b}\}$, or an outside alternative denoted by $(B, Y) = (0, 0)$. The set $\mathcal{M}_{BY} \equiv \{(b, y) : b \in \mathcal{B},\ y \in \mathcal{Y}_b\} \cup \{(0, 0)\}$ denotes the joint support of (B, Y). The set $\mathcal{Z}$ denotes the support of observable covariates Z, such as individual characteristics. The vector $V \equiv (V_1, ..., V_{\bar{b}})' \in \mathbb{R}^{\bar{b}}$ represents unobserved heterogeneity that affects individuals' preferences, where V_b denotes the b-th component of V. The distribution of V is denoted $G(\cdot)$ so that for any set $\mathcal{V} \subseteq \mathbb{R}^{\bar{b}}$, $G(\mathcal{V}) \equiv \mathbb{P}[V \in \mathcal{V}]$.

The utility obtained by an individual with covariates z and unobservable v from any choice $(b, y) \in \mathcal{M}_{BY}$ is given by

$$U_{by} \equiv u(b, y, z, v_b) \text{ if } (b, y) \neq (0, 0), \qquad U_{00} \equiv 0, \tag{2.1}$$

where $u(b, y, z, v_b)$ is strictly increasing in v_b for each (b, y, z) and the utility from the outside alternative is normalized to zero.

We assume that each individual chooses the alternative that maximizes her utility.[2] For any $b \in \mathcal{B}$ define $\bar{\mathcal{Y}}_b \equiv \mathcal{Y}_b \cup \{0\}$ and $U_{b0} \equiv 0$, and let

$$U_b^* \equiv \max_{y \in \bar{\mathcal{Y}}_b} U_{by}, \qquad Y_b^* \equiv \operatorname*{argmax}_{y \in \bar{\mathcal{Y}}_b} U_{by},$$

denote the indirect utility and optimal choice of Y, respectively, if the individual's alternatives were limited to only those of type b or the outside alternative. The structure of the model will be such that for any fixed b, the choice of the ordered outcome Y produces a standard model of ordered response, in the sense that this choice is weakly increasing in V_b. For example, if V_b is normally distributed, independent of Z, and the consumer may only purchase from brand b, then we have an ordered probit model. A consumer who has the option to choose any quality level from any brand then chooses

$$B = 1\left[\max_{b \in \mathcal{B}} U_b^* > 0\right] \cdot \operatorname*{argmax}_{b \in \mathcal{B}} U_b^*, \qquad Y = 1\left[\max_{b \in \mathcal{B}} U_b^* > 0\right] \cdot Y_B^*. \tag{2.2}$$

[2]Under Restriction A3 below ties in the utility obtained from different alternatives occur with zero probability conditional on any realization of z. How ties are handled is therefore of no consequence in the determination of conditional choice probabilities, but to simplify notation we adopt the convention that if alternatives (b, y) and (b, y'), $y < y'$, achieve the same utility, then (b, y) is chosen, and if $(b, y) \neq (b', y')$, $b < b'$ achieve the same utility, then (b', y') is chosen.

Restriction A1: (Probability space) (B, Y, Z, V) are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ contains the Borel sets. The support of (Z, V) is $\mathcal{Z} \times \mathbb{R}^{\bar{b}}$, where $\mathcal{Z}$ is a lattice, and the support of (B, Y) is $\mathcal{M}_{BY} \equiv \{(b, y) : b \in \mathcal{B},\ y \in \mathcal{Y}_b\} \cup \{(0, 0)\}$.

Restriction A2: (Identification of $f_z^0(b, y)$) For each value $z \in \mathcal{Z}$ there is a proper conditional distribution of (B, Y) given $Z = z$ and $f_z^0(b, y) \equiv \mathbb{P}[(B, Y) = (b, y)|Z = z]$ is point identified over the support of (B, Y) for almost every $z \in \mathcal{Z}$.

Restriction A3: (Distribution of unobserved heterogeneity) The distribution of V is absolutely continuous with respect to Lebesgue measure with everywhere positive density on $\mathbb{R}^{\bar{b}}$.

Restriction A4: (Independence) Z and V are stochastically independent.

Restriction A5: (Admissible structures) Structure $S \equiv (u, G)$ belongs to a known collection $\mathcal{S}$ of pairs of utility functions and distributions of unobserved heterogeneity, (u, G).

Restriction A6: (Utility maximization) Given (Z, V), (B, Y) are chosen to maximize $u(B, Y, Z, V_B)$, where u belongs to a known class of functions $\mathcal{U}$ satisfying (i) $u(0, 0, z, v_b) = 0$ for all (z, v_b), (ii) for all $(b, y) \neq (0, 0)$ and all z, $u(b, y, z, v_b)$ is strictly increasing and continuous in v_b, and (iii) for each $(b, z) \in \mathcal{B} \times \mathcal{Z}$, $\{u(b, y, z, v_b) : v_b \in \mathbb{R}\}$ satisfies the single-crossing property in (y, v_b), namely that if $v_b' > v_b$ and $y' > y$, then

$$u(b, y', z, v_b) - u(b, y, z, v_b) \geq (>)\, 0 \Rightarrow u(b, y', z, v_b') - u(b, y, z, v_b') \geq (>)\, 0.$$

Restrictions A1-A3 are standard. Restriction A1 defines the underlying probability space and notation for the support of random variables (B, Y, Z, V). Restriction A2 stipulates that the conditional distribution of (B, Y) given covariates z is point identified for almost every $z \in \mathcal{Z}$, as is the case for example under random sampling. Restriction A3 requires that unobserved heterogeneity V is absolutely continuously distributed with full support in Euclidean space.

Restriction A4 imposes independence of Z and V. This is an important restriction. If Z includes prices, then it requires that prices are exogenous, ruling out the possibility that unobserved components of individual utility are correlated with prices. This could be violated if different sellers offer different prices for the products being sold and if some individuals choose where to shop based on these prices. This assumption may still be appropriate however if the price of the product makes up only a small fraction of total expenditure – as is the case in our application to razor blade purchases – such that individuals are unlikely to choose where to shop for a basket of many goods (e.g. groceries and household products) based on this one product's price. In applications with individual level variation in which prices are thought to be endogenous, one could instead use an instrumental variable approach applicable to models of discrete choice such as those of Chesher, Rosen, and Smolinski (2013) and Chesher and Rosen (2017). This precludes the likelihood approach to inference employed here, resulting in moment inequality characterizations of identified sets, such as those used in applications to IV models for ordered response and models of interdependent binary outcomes in Chesher, Rosen, and Siddique (2019) and Chesher and Rosen (2020), respectively.

Restriction A5 defines a structure S as a utility function and distribution of unobserved heterogeneity, assumed to belong to some class of admissible pairs $\mathcal{S}$. Note that any given structure S gives rise to a collection of conditional distributions $f_z^0(b, y)$ for almost every $z \in \mathcal{Z}$. The identification problem is to determine the set of structures that can generate the observed distributions $f_z^0(b, y)$. The set of structures $\mathcal{S}$ admitted by the model can be restricted to a parametric, semiparametric, or nonparametric class.

The underlying structure S maps to conditional distributions $f_z^0(b, y)$ through the specification of the individual choice problem. Restriction A6 specifies that individuals choose (B, Y) to maximize utility $u(B, Y, Z, V_B)$, on which we impose some conditions. First, the specification (2.1) requires that there is a single, separate component of unobserved heterogeneity for each brand b, and through Restriction A6(ii) that utility from each product of this brand is weakly increasing in the associated unobservable. The components of V may however be jointly dependent, allowing for potential correlation across brand preferences, and quality tastes across brands. With Restriction A6(i) we normalize the utility from the outside option to zero. Restriction A6(iii) requires that the utility function satisfies the single-crossing property in (y, v_b). By Milgrom and Shannon (1994) Theorem 4 this guarantees that for all consumers and all $b \in \mathcal{B}$, the optimal choice within brand b, Y_b^*, is nondecreasing in v_b, so that quality choice within any brand b assumes the structure of an ordered choice problem. Thus, this restriction ensures the alternatives are partially ordered, because there is an ordered response structure to the choice problem within (but not across) brands. The following Section provides a characterization of conditional choice probabilities and subsequent identification analysis under these restrictions.

3 Identification

We begin this section by characterizing in Section 3.1 the form of the multivariate integral delivering conditional choice probabilities as a function of the underlying structure S. In Section 3.2 we then provide a general characterization of the identified set of structures compatible with Restrictions A1-A6, and we show that if the model is correctly specified, the identified set can be written as the maximizers of the expected log-likelihood. In Section 4 we then provide specialized identification results for a class of parametric models.

3.1 Conditional Choice Probabilities

The utility maximization hypothesis together with the shape restrictions in Restriction A6 enable concise characterization of the conditional choice probabilities

$$\wp_{by}(z; S) \equiv G(\mathcal{V}_{by}(z; u))$$

for brand-quality pair (b, y) given $Z = z$, considered as a function of any structure $S = (u, G)$, where

$$\mathcal{V}_{by}(z; u) \equiv \left\{ v \in \mathcal{V} : \forall \left(\tilde{b}, \tilde{y}\right) \neq (b, y), \quad u(b, y, z, v_b) \geq u\left(\tilde{b}, \tilde{y}, z, v_{\tilde{b}}\right) \right\} \tag{3.1}$$

denotes the set of values for unobserved heterogeneity V on which (b, y) maximizes utility given z and v. Without parametric restrictions on u, the monotonicity and single-crossing conditions suffice to establish the representation of each choice probability $\wp_{by}(z; S)$ as a particular form of a $\bar{b}$-variate integral. Thus, given a specific (u, G), $\wp_{by}(z; S)$ can be computed by either numerical integration or simulation. The formal result follows.

Theorem 1 *Let Restriction A6 hold. Then for each $(b, y, z) \in \mathcal{M}_{BY} \times \mathcal{Z}$, the region $\mathcal{V}_{by}(z; u)$ is a convex polytope in $\mathbb{R}^{\bar{b}}$ defined by the inequalities*

$$V_b \in (g_b(y), g_b(y+1)], \tag{3.2}$$

$$\forall k < b,\ V_k \leq h_{b,k}(y), \text{ and} \tag{3.3}$$

$$\forall k > b,\ V_k < h_{b,k}(y), \tag{3.4}$$

and additionally under Restriction A4 the conditional choice probability $\wp_{by}(z; S)$ takes the *form*

$$\wp_{by}(z;S) = \int_{g_b(y)}^{g_b(y+1)} \left(\int_{H(b,y)} dG_{V_{-b}|V_b}(v_{-b}|v_b) \right) dG_{V_b}(v_b), \tag{3.5}$$

where

$$H(b,y) \equiv \{r \in \mathbb{R}^{\bar{b}-1} : r_1 \leq h_{b,1}(y), \ldots, r_{b-1} \leq h_{b,b-1}(y), r_b \leq h_{b,b+1}(y), \ldots, r_{\bar{b}-1} \leq h_{b,\bar{b}}(y)\}$$

with $\{h_{b,k}(y) : k \neq b\}$ *cross-brand threshold functions, and* $\{g_b(y) : y = (0, ..., \bar{y}_b + 1)\}$ *within-brand threshold functions satisfying* $g_b(0) \equiv -\infty$ *and* $g_b(\bar{y}_b + 1) \equiv \infty$*. Here* $G_{V_{-b}}(\cdot|v_b)$ *and* $G_{V_b}(\cdot)$ *denote the conditional distribution of* V_{-b} *given* $V_b = v_b$ *and the marginal distribution of* V_b*, respectively. The threshold function* $g_b(\cdot)$ *may depend on* z *and each function* $h_{b,k}(\cdot) : k \neq b$*, may depend on both* v_b *and* z*.*

3.2 General Characterization of the Identified Set

Before adding further restrictions we first characterize the identified set of structures S under Restrictions A1-A6, denoted $\mathcal{S}_0$ and defined as

$$\mathcal{S}_0 \equiv \left\{(u,G) \in \mathcal{S} : \forall (b,y) \in \mathcal{M}_{BY},\, G(\mathcal{V}_{by}(z;u)) = f_z^0(b,y) \text{ a.e. } z \in \mathcal{Z}\right\}, \tag{3.6}$$

where $\mathcal{V}_{by}(z;u)$ denotes the set of values of V defined by (3.2)–(3.4) in Theorem 1. In words, $\mathcal{S}_0$ is the set of admissible structures (u,G) that generate identified conditional choice probabilities $f_z^0(b,y)$ for all (b,y) and almost every $z \in \mathcal{Z}$. Given the absolute continuity of the distribution of V and the continuity of utility in unobserved heterogeneity, the intersection of sets $\mathcal{V}_{by}(z;u)$ and $\mathcal{V}_{\tilde{b}\tilde{y}}(z;u)$, $\left(\tilde{b},\tilde{y}\right) \neq (b,y)$, has Lebesgue measure zero, so that there is a unique utility maximizing pair (b,y) with probability one given any $z \in \mathcal{Z}$. Hence $G(\mathcal{V}_{by}(z;u))$ is the conditional probability of observing (b,y) given $Z = z$ when the utility function is u and $V \sim G$. Structures (u,G) that do not belong in the identified set $\mathcal{S}_0$ in (3.6) are those such that the set

$$\mathcal{Z}^*(u,G) \equiv \left\{z \in \mathcal{Z} : \exists (b,y) \in \mathcal{M}_{BY} \text{ s.t. } G(\mathcal{V}_{by}(z;u)) \neq f_z^0(b,y)\right\}, \tag{3.7}$$

has positive measure $\mathbb{P}_Z$.

Given the representation of the identified set through the equalities $G(\mathcal{V}_{by}(z;u)) = f_z^0(b,y)$ we can equivalently characterize the identified set as those structures that maximize the log-likelihood. For this we require that the model is correctly specified, formalized with the following additional assumption.

Restriction A7: (Correct Specification) $\exists S^* \in \mathcal{S}$, $S^* \equiv (u^*, G^*)$ such that $\forall (b,y) \in \mathcal{M}_{BY}$ $G^*(\mathcal{V}_{by}(z;u^*)) = f_z^0(b,y)$ a.e. $z \in \mathcal{Z}$.

Consider the expected log-likelihood function

$$Q(u,G) \equiv E[\ln G(\mathcal{V}_{BY}(Z;u))],$$

where the expectation is taken with respect to population measure $\mathbb{P}$. It follows by arguments identical to those with singleton $\mathcal{S}_0$ that $Q(u,G)$ attains its maximum at all $(u,G) \in \mathcal{S}_0$, since by definition these all produce the same probabilities $G(\mathcal{V}_{by}(z;u))$ for almost every z. The general observation that when point identification is lacking the set of maximizers of the expected log-likelihood are precisely those observationally equivalent to the population data generating structure has been made previously, see e.g. Bowden (1973) and Redner (1981). The formal statement in the present setting, a proof of which is included in the Appendix for completeness, is made in the following Proposition.

Proposition 1 *Let restrictions A1-A7 hold. Then* $\mathcal{S}_0 = \underset{(u,G)\in\mathcal{S}}{\operatorname{argmax}}\, Q(u,G)$.

Unless sufficiently strong parametric restrictions on $\mathcal{S}$ are imposed, $\mathcal{S}_0$ may not be a singleton, so that there may not be point identification. When sufficiently strong restrictions for point identification do hold, estimation and inference can proceed under the classical maximum likelihood paradigm. When these restrictions do not hold, the classical results do not apply. But the characterization of $\mathcal{S}_0$ as the (set of) maximizers of the expected log-likelihood enables us to apply inference techniques for maximum likelihood estimators when point identification is lacking. The subsequent characterization of choice probabilities $G(\mathcal{V}_{by}(z;u))$ under parametric restrictions facilitates derivation of some sufficient conditions for point identification, as well as computation of set estimates and inferential statistics when point identification fails.

4 Identification under Parametric Restrictions

In this section we consider a parametric restriction on the underlying utility functions with $\bar{b}$ firms $b \in \mathcal{B} \equiv \{1,2,...,\bar{b}\}$, each selling a low-quality product offering $(Y=1)$ and a high-quality product offering ($Y = 2$), so that for all $b \in \mathcal{B}$, $\mathcal{Y}_b \equiv \{1,2\}$.[3] We denote covariates $Z \equiv (Z_1, ..., Z_{\bar{b}})$ with $Z_b \equiv (X_b, P_b)$ comprising covariates that affect the utility of choosing a brand b offering, allowing for the possibility that Z_b and $Z_{\tilde{b}}$, $b \neq \tilde{b}$, contain common components. We specify

$$u\left(b, y, z, v_b\right) \equiv y \times \left(x_b \beta_b + v_b\right) - \alpha_{by}, \qquad \alpha_{00} \equiv 0, \tag{4.1}$$

for each $(b,y) \in \mathcal{M}_{BY}$, $z \in \mathcal{Z}$, and $v_b \in \mathbb{R}$. Realizations of covariates X_b contribute to the linear index $x_b\beta_b + v_b$, while P_b are covariates that may enter in the determination of α_{by}. The specification of $x_b\beta_b + v_b$ entering utility multiplicatively in y is a parametric restriction that results in the optimal choice restricted to any fixed $b \in \mathcal{B}$ taking a parametric linear index form familiar from parametric ordered response models, such as ordered probit and logit. The parameters α_{by} capture that component of utility from choice (b, y) not restricted to scale linearly with y. These in turn determine thresholds λ_{by} for individuals' within-brand ordered choice preferences as described in (4.2) below. We consider variants of our model in which α_{by} are fixed parameters for each (b, y), as well as cases where they may be parametrically specified functions of observable covariates, such as prices. This will lead to generalizations of ordered choice models in which threshold parameters may or may not depend on observable exogenous variables.

This model generalizes a three-choice ordered response model, such as ordered probit or logit, in that for any fixed $b \in \mathcal{B}$ we have

$$\begin{aligned} Y_b^* &= 0 \Leftrightarrow \lambda_{b0} < V_b \leq \lambda_{b1} - X_b\beta_b, \\ Y_b^* &= 1 \Leftrightarrow \lambda_{b1} - X_b\beta_b < V_b \leq \lambda_{b2} - X_b\beta_b, \\ Y_b^* &= 2 \Leftrightarrow \lambda_{b2} - X_b\beta_b < V_b, \end{aligned} \tag{4.2}$$

where

$$\lambda_{b0} \equiv -\infty, \quad \lambda_{b1} \equiv \min\left\{\alpha_{b1}, \frac{\alpha_{b2}}{2}\right\}, \quad \lambda_{b2} \equiv \max\left\{\alpha_{b2} - \alpha_{b1}, \frac{\alpha_{b2}}{2}\right\}. \tag{4.3}$$

denote threshold parameters. Some algebra reveals that

$$\alpha_{b2} > 2\alpha_{b1} \Rightarrow \lambda_{b1} = \alpha_{b1}, \text{ and } \lambda_{b2} = \alpha_{b2} - \alpha_{b1}, \tag{4.4}$$

[3]Extension of the analysis to the case where firms sell more than two product offerings is conceptually straightforward. We focus on the case where each brand has two ordered product offerings to economize on notation and also because this is the setting encountered in our application.

while

$$\alpha_{b2} \leq 2\alpha_{b1} \Rightarrow \lambda_{b1} = \lambda_{b2} = \frac{\alpha_{b2}}{2}. \tag{4.5}$$

The inequality on the left hand side of (4.4) ensures that for each b,

$$\mathbb{P}\left[\alpha_{b1} - X_b\beta_b < V_b \leq \alpha_{b2} - \alpha_{b1} - X_b\beta_b | X = x\right] > 0,$$

or equivalently that some randomly chosen individuals prefer $y = 1$ to both the other alternative of type b and the outside alternative. When instead the inequality on the left hand side of (4.5) holds, then the probability of this event is zero. In this case, if one were to imagine taking a randomly selected individual and increasing their unobservable V_b continuously from $-\infty$ to ∞, that individual would choose the outside alternative for values of V_b up to $\frac{\alpha_{b2}}{2} - X_b\beta_b$, and then switch to $Y = 2$ for all $V_b > \frac{\alpha_{b2}}{2} - X_b\beta_b$, respecting the ordered nature of the quality dimension y, but skipping over the lower quality alternative $y = 1$.

Given the parametric specification (4.1) for u, the resulting regions of unobserved variables $\mathcal{V}_{by}$ defined by (3.2)–(3.4) in Theorem 1 take the form of convex polytopes in $\mathbb{R}^{\bar{b}}$. Figure 1 provides an example illustrating these regions for a case in which $\bar{b} = 2$ for given values of $\beta_1, \beta_2, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}$, and a given value of the conditioning variables z in which the inequality $\alpha_{b2} > 2\alpha_{b1}$ on the left hand side of (4.4) holds.

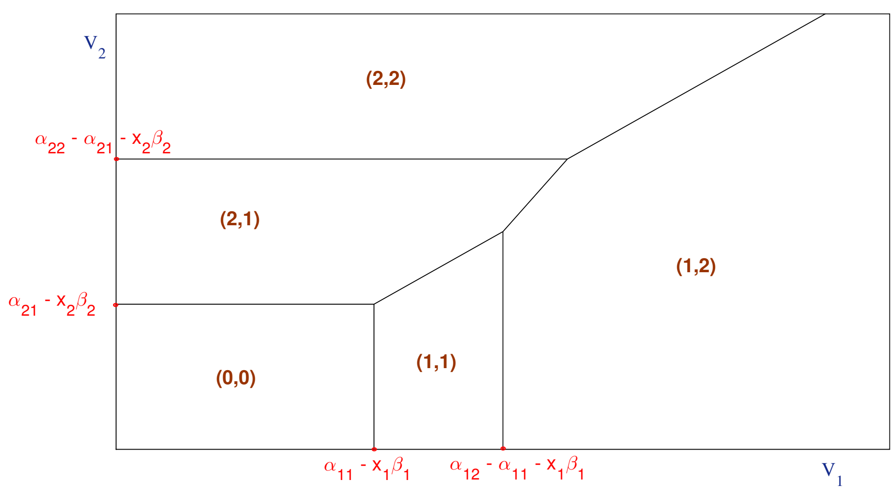

Figure 1: Regions of unobservables V resulting in each choice of $(b, y) \in \mathcal{M}_{BY}$ with utility as specified in (4.1).

The specification considered so far has implicitly treated $(\alpha_{11}, \alpha_{12}, ..., \alpha_{\bar{b}1}, \alpha_{\bar{b}2})$ as fixed parameters to be estimated. Fixed threshold specifications for ordered response models are common, but it is straightforward to allow thresholds to be functions of observable variables P_b. This is important in our application, where observed prices may affect the utility of purchasing each product.

Here and in our application for each $b \in \mathcal{B}$ and $y \in \{1,2\}$, α_{by} is specified by

$$\alpha_{by} = \delta_b + k\left(p_{by}, \gamma_b\right), \qquad k\left(p_{by}, \gamma_b\right) = \gamma_b p_{by}, \tag{4.6}$$

where the observed variable $p_{by} \in \mathbb{R}$ denotes the price of alternative (b, y), and δ_b and γ_b denote parameters on the real line. More generally, p_{by} could be used to denote any vector of observable variables thought to influence α_{by}. As before the utility of the outside alternative is normalized to zero. The optimal choice Y_b^* for fixed brand b is still determined by (4.2) and (4.3), but with each α_{by} determined by (4.6).[4]

With price now entering the utility function, the possibility of values of exogenous variables that imply that consumers never choose quality offering $Y = 1$ for a given brand $B = b$ becomes particularly relevant. This is because unlike covariates X_b, brand b prices $P_b \equiv (P_{b1}, P_{b2})$ vary across the vertical dimension y within brand.[5] In contrast to a fixed threshold specification, it is possible that there are values of the conditioning variables $Z = z$ such that the conditional choice probability $\mathbb{P}\left[(B, Y) = (b, 1) \,|\, Z = z\right]$ equals zero for either b, while conditional on other values $Z = \tilde{z}$, $\mathbb{P}\left[(B, Y) = (b, 1) \,|\, Z = \tilde{z}\right] > 0$. This is practically relevant because there may be consumers who face prices such that the higher quality product offering will always be more desirable than the lower product quality offering no matter their realization of unobservables V, as could happen when a firm introduces a sale for the high quality offering in order to induce consumers to try it. Thus both cases (4.4) and (4.5) are allowed in all that follows, depending on the value of conditioning variables Z.

We now specialize the characterization of the set $\mathcal{S}_0$ from Proposition 1 and the form of the conditional choice probabilities given in Theorem 1 to models in which utilities satisfy (4.1) such that formally $\mathcal{S} \equiv \mathcal{U} \times \mathcal{G}$, where

$$\mathcal{U} \equiv \left\{ \begin{array}{c} u : \mathcal{B} \times \mathcal{Y} \times \mathcal{Z} \times \mathbb{R}^{\bar{b}} \to \mathbb{R} : \; u\left(b, y, z, v; \theta\right) \equiv y\left(x_b \beta_b + v_b\right) - \delta_b - \gamma_b p_{by} \\ \text{for some } \theta \equiv \left\{\left(\delta_b, \gamma_b, \beta_b\right) : b \in \mathcal{B}\right\} \in \Theta. \end{array} \right\}, \tag{4.7}$$

[4]We focus here on a linear specification for $k\left(p_{by}, \gamma_b\right)$, but alternative specifications are possible. For instance, one could use the CRRA or isoelastic utility specification $k\left(p_{by}, \gamma_b\right) = \left(1 - \gamma_b\right)^{-1} p_{by}^{1-\gamma_b}$ with $\gamma_b > 0$, or the exponential utility specification $k\left(p_{by}, \gamma_b\right) = 1 - \exp\left(-\gamma_b p_{by}\right)$ for some $\gamma_b \in \mathbb{R}$.

[5]Following the notation introduced at the beginning of this section, covariates that affect the utility from brand b offerings are thus $Z_b \equiv (X_b, P_{b1}, P_{b2})$, where $Z \equiv (Z_1, ..., Z_b)$.

where Θ is a compact subset of Euclidean space and $\mathcal{G}$ denotes a collection of distribution functions for V indexed by parameters Σ, belonging to compact parameter space $\mathbf{\Lambda}$. We define $\zeta \equiv (\theta, \Sigma)$ to denote the full parameter vector. The utility function and distribution pair (u, G) is completely specified given ζ, so under this parametric specification we use ζ to denote the corresponding structure $(u, G) \in \mathcal{S}$, with $G(\cdot; \Sigma)$ denoting the distribution G indicated by parameter Σ. Then by definition the identified set for ζ is

$$\mathcal{S}_0 \equiv \left\{ \zeta \in \Theta \times \mathbf{\Lambda} : \forall (b, y) \in \mathcal{M}_{BY},\ \wp_{by}(z; \zeta) = f_z^0(b, y) \text{ a.e. } z \in \mathcal{Z} \right\},$$

where $f_z^0(b, y)$ is the population probability that $B = b$ and $Y = y$ conditional on $Z = z$.

Application of Proposition 1 gives the likelihood characterization of the identified set

$$\mathcal{S}_0 \equiv \underset{\zeta \in \Theta \times \mathbf{\Lambda}}{\operatorname{argmax}}\ E\left[\ln \wp_{BY}(Z; \zeta)\right] = \underset{\zeta \in \Theta \times \mathbf{\Lambda}}{\operatorname{argmax}}\ E_z E\left[\ln \wp_{BY}(Z; \zeta) | Z = z\right].$$

Using the parametric structure set out above together with Theorem 1 we have

$$\wp_{by}(z; \zeta) = \int_{g_b(y; z, \theta)}^{g_b(y+1; z, \theta)} \left(\int_{H(b, y, z, v_b, \theta)} dG_{V_{-b}|V_b = v_b}(v_{-b}; \Sigma) \right) dG_{V_b}(v_b; \Sigma),$$

where

$$g_b(y; z, \theta) \equiv \lambda_{by} - X_b \beta_b, \tag{4.8}$$

with each λ_{by} as defined by (4.3) and (4.6), and where the region of integration $H(b, y, z, v_b, \theta)$ in the inner integral is the set of $r \in \mathbb{R}^{\bar{b}-1}$ such that

$$r_1 \leq h_{b,1}(y, z, v_b, \theta), ..., r_{b-1} \leq h_{b,b-1}(y, z, v_b, \theta), r_b \leq h_{b,b+1}(y, z, v_b, \theta), ..., r_{\bar{b}-1} \leq h_{b,\bar{b}}(y, z, v_b, \theta),$$

where for all $d \neq b$ and $v \in \mathbb{R}$,

$$h_{b,d}(y, z, v, \theta) \equiv \min_{\tilde{y} \in \{1, \ldots, \bar{y}_d\}} \frac{1}{\tilde{y}} \left[y \left(x_b \beta_b + v \right) - \left(\alpha_{by} - \alpha_{d\tilde{y}} \right) \right] - x_d \beta_d, \tag{4.9}$$

with α_{by} as specified in (4.6). Thus each $\wp_{by}(z; \zeta)$ takes the form of an integral over a region defined by inequalities that are linear in θ. Written in this form it is straightforward to verify that $\wp_{by}(z; \zeta) = \wp_{by}(z; (\theta, \Sigma))$ is log-concave in θ for any fixed Σ and each value (b, y, z) if G is a known log-concave distribution with density $f_V(\cdot, \Sigma)$. This in turn implies that the maximizers of $\mathcal{L}(\zeta)$ for any fixed Σ comprise a convex set.

Theorem 2 *Suppose that Restrictions A1-A7 hold, that* $u \in \mathcal{U}$ *defined in (4.7), and that* $G(\cdot) = G(\cdot;\Sigma)$ *with known* Σ *such that* $G(\cdot;\Sigma)$ *has log-concave density* $f_V(\cdot,\Sigma)$*. Then the identified set for* θ *is*

$$\Theta^*(\Sigma) \equiv \underset{\theta\in\Theta}{\operatorname{argmax}}\ \mathcal{L}(\theta,\Sigma),$$

with the expected log-likelihood

$$\mathcal{L}(\theta,\Sigma) \equiv \sum_{(b,y)\in\mathcal{M}_{BY}} f_z^0(b,y)\ln \int_{\mathbb{R}^{\bar{b}}} f_V(v,\Sigma)\, 1\left[v \in \mathcal{V}_{by}(z;\theta)\right] dv,$$

concave in θ*.*

Many commonly used distributions are log-concave, with the multivariate normal distribution a leading example. If the distribution G is not known, but the elements of the admissible set of distributions $\mathcal{G}$ are all log-concave, for example if all such distributions are multivariate normal but with different variances, then it follows that the identified set for θ is contained in a union of convex sets, namely the union of sets delivered by Theorem 2 for each $G \in \mathcal{G}$.

Under some additional conditions on the variation in observable variables Z, a known G can deliver point identification, as stated in Theorem 3 below. The Theorem is a generalization of a result in Theorem 2 of Aradillas-Lopez and Rosen (2022), up to minor changes in notation, allowing for $\bar{b} > 2$ and also focusing on regions of the parameter space in which for all b $\alpha_{b2} > 2\alpha_{b1}$, equivalently $\wp_{b1}(z;\zeta) > 0$. Accordingly, define this region as

$$\mathcal{Z}^* \equiv \{z \in \mathcal{Z} : \forall b \in \mathcal{B},\ \wp_{b1}(z;\zeta) > 0\}.$$

The reason the result from Aradillas-Lopez and Rosen (2022) applies with $\bar{b} = 2$ is the equivalence of the conditional probability of consumers choosing not to purchase, i.e. $\wp_0(z;\zeta)$ in the present model, to the conditional probability that $(0,0)$ is an equilibrium in the ordered outcome simultaneous equations model studied by Aradillas-Lopez and Rosen (2022). While both models feature the same conditional probabilities for these outcomes, the rest of their observable implications differ. The simultaneous equations model of Aradillas-Lopez and Rosen (2022) produces *inequalities* on the conditional probabilities of other outcomes, due to the presence of strategic interactions and multiple equilibria. They then combine the conditional moment equality from the probability of outcome $(0,0)$ with conditional moment inequalities to produce a test statistic for inference. In the single agent decision problem studied here, the model delivers *equalities* for the conditional probabilities of *all* outcomes. The formal result is now provided, wherein Θ_b is used to denote the projection of the parameter space for θ onto the space of admissible $(\delta_b, \gamma_b, \beta_b)$.

Theorem 3 *Suppose that Restrictions A1-A7 hold, that $\mathbb{P}\{Z \in \mathcal{Z}^*\} > 0$, and that we have the parametric structure $\mathcal{S} = \mathcal{U} \times \mathcal{G}$ given in (4.7) with singleton $\mathcal{G}$ so that Σ is known. Then if (i) conditional on $Z \in \mathcal{Z}^*$ for each $b \in \mathcal{B}$ there exists no proper linear subspace of the support of $\tilde{Z}_b \equiv (1, P_{b1}, -X_b)$ that contains $\tilde{Z}_b$ with probability one, and (ii) for all conformable column vectors $c_1, ..., c_{\bar{b}}$ satisfying $c_b \in \{\theta_b - \tilde{\theta}_b : \theta_b \in \Theta_b,\ \tilde{\theta}_b \in \Theta_b\}$ for each b, with $c_b \neq 0$ for some b, we have that at least one of*

$$\mathbb{P}\left\{\tilde{Z}_1 c_1 \leq 0, \ldots, \tilde{Z}_{\bar{b}} c_{\bar{b}} \leq 0 | \tilde{Z}_b c_b < 0, Z \in \mathcal{Z}^*\right\} > 0 \text{ with } \mathbb{P}\left\{\tilde{Z}_b c_b < 0, Z \in \mathcal{Z}^*\right\} > 0, \quad (4.10)$$

or

$$\mathbb{P}\left\{\tilde{Z}_1 c_1 \geq 0, ..., \tilde{Z}_{\bar{b}} c_{\bar{b}} \geq 0 | \tilde{Z}_b c_b > 0, Z \in \mathcal{Z}^*\right\} > 0 \text{ with } \mathbb{P}\left\{\tilde{Z}_b c_b > 0, Z \in \mathcal{Z}^*\right\} > 0 \quad (4.11)$$

holds then θ is point identified.

The theorem above shows that under conditions that guarantee sufficient variation in exogenous variables, θ is point identified. The first requirement is that $\mathbb{P}\{Z \in \mathcal{Z}^*\} > 0$, i.e. that there is positive probability of values of exogenous variables such that $\mathbb{P}[B = b, Y = 1 | Z = z] > 0$ for all b. The subsequent statements (i) and (ii) are then made conditional on $Z \in \mathcal{Z}^*$. Condition (i), is standard. Note that this requires that each X_b contains no constant components. Condition (ii) implies that conditional on $\tilde{Z}_b c_b$ negative (positive), each $\tilde{Z}_d c_d$, $d \neq b$ takes nonpositive (nonnegative) values with nonzero probability. This condition helps to achieve identification because it ensures that for any $\tilde{\theta} \neq \theta$ there exist values of z such that as a function of $\tilde{\theta}$ the indices that define the cutoffs for the outside option are all either above or below the corresponding indices at θ, with the comparison being strict for at least one index. This implies that the implied conditional probabilities of choosing the outside alternative for the two values $\tilde{\theta}$ and θ differ for such values of Z.

The support requirement of condition (ii) will in general depend on the specification of the parameter space Θ. If the parameter space is an arbitrary compact subset of Euclidean space with no restrictions imposed on $\theta_b = (\delta_b, \gamma_b, \beta_b)$ across b it will hold under standard large support assumptions requiring each Z_b to have a component taking arbitrarily large values with positive probability conditional on $Z \in \mathcal{Z}^*$. In this case it appears difficult to come up with general weaker support conditions that will imply (ii). However, in models in which there are restrictions involving both θ_b and $\theta_{b'}$, $b \neq b'$, weaker conditions can suffice. For example, if the parameters θ_b relevant to the $\mathcal{V}_{00}(z;\theta)$ region are restricted to be the same for all b, as is the case in the application of Aradillas-Lopez and Rosen (2022) in which $\bar{b} = 2$, then this condition can be satisfied even if each $\tilde{Z}_b$ has a discrete distribution. In this case $\theta_b = \theta$ for all b, so we can also write $c_b = c$ for each b. Condition (4.10) then becomes $\mathbb{P}\left\{\tilde{Z}_1 c \leq 0, ..., \tilde{Z}_{\bar{b}} c \leq 0 | \tilde{Z}_b c < 0, Z \in \mathcal{Z}^*\right\} > 0$, which is easier to satisfy. It is trivial for instance if $\tilde{Z}_b = \tilde{Z}$ for all b.

Notably, the theorem hinges only on implied differences in the conditional probability of $Y = 0$ given Z, and exploits no additional information from the conditional probabilities of other brand-quality combinations. The conditions are shown to be sufficient, but are not shown to be necessary, so point identification may hold under weaker conditions.

Although the conditions of Theorem 3 are not directly applicable to specifications in which G is restricted to belong to a family of admissible distributions $\mathcal{G}$, it still carries meaningful implications in such contexts, and it is useful for understanding what type of variation is helpful for identification. Under the other conditions of the Theorem, associated with each possible $G' \in \mathcal{G}$ there can be only a singleton identified set for θ, say $\theta(G')$. Thus the identified set can only consist of parameter values for θ that are $\theta(G')$ *for some* $G' \in \mathcal{G}$. This implies that with a parametric specification in which $\mathcal{G}$ is parameterized by Σ, each element of the identified set for θ must maximize $\mathcal{L}(\zeta) = \mathcal{L}(\theta, \Sigma)$ for some Σ.

A leading example of a model for which this Theorem is applicable is a partially ordered logit model in which unobservable variables $V_1, ..., V_{\bar{b}}$ are assumed to be i.i.d. logit variates. This restriction would simplify computation, but at the cost of imposing that unobservable taste for the vertical dimension is independent of unobservable taste in the horizontal dimension. While this may be reasonable in some settings, in our application this would require unobserved preference for quality to be independent of brand preferences. Thus, we employ a specification in which the unobservables are restricted to be multivariate normal, with variance governed by parameters to be estimated. When allowance for potential correlation among unobservables is desired other parametric specifications for their joint distribution could also be used, such as the Farlie-Gumbel-Morgenstern copula in conjunction with logit marginals employed by Aradillas-Lopez and Rosen (2022).[6]

[6]This would retain the convenience of logit marginals while allowing unobservables to be correlated, but

5 CCPs for the Partially Ordered Probit Model

In this section ζ will be used to denote the full vector of model parameters of a bivariate probit model with $\bar{b} = \bar{y} = 2$ and utility function as specified by (4.7), and $G(\cdot;\Sigma)$ corresponding to a bivariate normal distribution with variance matrix Σ with Σ_{11} normalized to one, $\Sigma_{22} = \sigma$ and $\Sigma_{12} = \Sigma_{21} = \rho\sigma$, where ρ, σ are unknown parameters. Thus $\zeta \equiv (\gamma_1, \gamma_2, \delta_1, \delta_2, \beta_1, \beta_2, \rho, \sigma)$, where each β_b is a vector of coefficients on variables X_b that affect utility from alternatives from brand b. This is the specification used in our application in Section 6. The parameter space for ζ is denoted Υ, the parameter space for ζ_k is denoted Υ_k, and Υ coincides with the product of Υ_k across $k = 1, ..., \dim(\zeta)$. We focus attention on the partially ordered probit model, but other specifications for the distribution of unobservable heterogeneity could be used as alternatives, such as the partially ordered logit specification discussed at the end of the previous section.

The choice probabilities $\wp_{by}(z, \zeta)$ implied by the partially ordered probit model must be computed in order to compute the likelihood. These are of the form set out in (3.5), which with $\bar{b} = 2$ take the form of a bivariate integral and can thus be computed using numerical integration or simulation for any given (ζ, z_i). Although we experimented with implementing both approaches, maximization of the log-likelihood was found to perform relatively slowly using these methods. With (ρ, σ) unknown the log-likelihood is generally not concave in parameters.

To compute the choice probabilities (and therefore the log-likelihood) more quickly, we used results from Owen (1980) that allow us to show equivalence of the choice probabilities to a closed form expression that does not involve integration. This alternative formulation involves univariate and bivariate normal CDFs evaluated at functions of parameters and observable variables. Software was used that vectorizes function evaluation to compute the CDF at each value of z_i in one function call.[7] This enabled computing likelihood contributions significantly more quickly than performing numerical integration or computing simulated probabilities separately for each observation. The details of how the conditional choice probabilities were manipulated to bypass the need for explicitly computing or simulating integrals are now set out. Section 6.2 then explains how inference was implemented. The finite sample performance of inference methods using both Wald and quasi-likelihood ratio (QLR) statistics is investigated in Monte Carlo experiments reported in Appendix D.

---

only to a limited extent relative to the bivariate normal specification employed in our application.

[7] We used the pbivnorm R package Kenkel (2015), which is based on Azzalini and Genz (2016).

In the partially ordered probit model in which $G(\cdot;\Sigma)$ is the bivariate normal distribution, application of (3.5) gives the following representation for the conditional choice probabilities:

$$\forall (b,y) \in \mathcal{M}_{BY}, \qquad \wp_{by}(z,\zeta) = \frac{1}{\sigma_b} \int_{g_b(y;z,\theta)}^{g_b(y+1;z,\theta)} \Phi\left(\frac{h_b(y,z,v,\theta) - \rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv, \quad (5.1)$$

$$h_b(y,z,v,\theta) \equiv \min_{\tilde{y}\in\{1,\ldots,\bar{y}_d\}} \frac{1}{\tilde{y}}\left[y\left(x_b\beta_b + v\right) - \left(\alpha_{by} - \alpha_{d\tilde{y}}\right)\right] - x_d\beta_d, \quad (5.2)$$

where $d = 3 - b$ denotes the brand other than b.[8]

To remove the need to simulate or numerically approximate the above integral, conditional choice probabilities $\wp_{by}(z,\zeta)$ can be further simplified using formulas for integrals of normal densities and distribution functions collected in Owen (1980). The representation so obtained is given in the following Proposition.

Proposition 2 *Let Restrictions A1-A7 hold with* $\bar{b} = 2$, $\bar{y}_b = 2$ *for each* b, *and*

$$u(b,y,z,v_b) \equiv y \times (x_b\beta_b + v_b) - \alpha_{by},$$

as in (4.1) with $\alpha_{00} \equiv 0$ *and* $V = (V_1, V_2)$ *normally distributed with mean zero and variance matrix* $\Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}$ *with unknown parameters* $\rho \in [-1,1]$ *and* $\sigma > 0$. *Then the conditional choice probabilities for each* $b = 1,2$ *and* $y = 1,2$ *can be expressed as*

$$\wp_{by}(z,\zeta) = \begin{pmatrix} 1\left[z^*_{by} < g_b(y+1;z,\theta)\right]\Delta\left(\sigma_b^{-1}\max\left\{z^*_{by}, g_b(y;z,\theta)\right\}, \sigma_b^{-1}g_b(y+1;z,\theta), m_1^+, m_2^+\right) \\ +1\left[z^*_{by} > g_b(y;z,\theta)\right]\Delta\left(\sigma_b^{-1}g_b(y;z,\theta), \sigma_b^{-1}\min\left\{z^*_{by}, g_b(y+1;z,\theta)\right\}, m_1^-, m_2^-\right) \end{pmatrix}, \quad (5.3)$$

where $g_b(\cdot;z,\theta)$ *is as defined in (4.8) and for any* h, k, m_1, m_2,

$$\Delta(h,k,m_1,m_2) \equiv \Phi_2(k,m_1;m_2) - \Phi_2(h,m_1;m_2), \quad (5.4)$$

where $\Phi_2(\cdot,\cdot,\rho)$ *denotes the CDF of a bivariate normal random vector* Z *with mean zero and unit variance components with correlation* ρ, *and for* $d = 3 - b$,

$$z^*_{by} \equiv \frac{\alpha_{d2} + \alpha_{by} - 2\alpha_{d1}}{y} - x_b\beta_b,$$

[8] Here $h_b(y,z,v,\theta)$ is equal to $h_{b,d}(y,z,v,\theta)$ defined in (4.9) with $d = 3 - b$ since there are only two firms.

and

$$m_1^+ \equiv \frac{yx_b\beta_b + \alpha_{d2} - \alpha_{by} - 2x_d\beta_d}{\sqrt{\sigma_b^2 y^2 - 4\rho\sigma_b\sigma_d y + 4\sigma_d^2}}, \quad m_2^+ \equiv \frac{2\rho\sigma_d - \sigma_b y}{\sqrt{\sigma_b^2 y^2 - 4\rho\sigma_b\sigma_d y + 4\sigma_d^2}}, \tag{5.5}$$

$$m_1^- \equiv \frac{yx_b\beta_b + \alpha_{d1} - \alpha_{by} - x_d\beta_d}{\sqrt{\sigma_b^2 y^2 - 2\rho\sigma_b\sigma_d y + \sigma_d^2}}, \quad m_2^- \equiv \frac{\rho\sigma_d - \sigma_b y}{\sqrt{\sigma_b^2 y^2 - 2\rho\sigma_b\sigma_d y + \sigma_d^2}}. \tag{5.6}$$

6 Application to Razor Blade Purchases

This section presents an application of the parametric bivariate probit model in Section 5 to the market for women's razor blades using consumer data from Great Britain. We discuss the data, computational details, and empirical results in turn.

6.1 Data

We use purchase data for a rolling panel of households from the Kantar FMCG Purchase Panel. The data comprise a representative sample of households observed making repeated purchases, obtained by a handheld scanner used to record all household grocery purchases at the UPC level. Data on razor blade purchases is used for the years 2004 – 2005.[9] In particular we focus on consumers' decisions to buy a double or triple blade cartridge from one of the two leading razor blade brands, Gillette and Wilkinson Sword.

In our application, we consider households observed purchasing either razor blade cartridges for a reusable non-electric women's razor (which we refer to as "system blades") or disposable women's razors, where the main shopper of the household is a female between the age of 18 and 50 years old. The outside alternative is the purchase of a disposable razor. The total sample size consists of 4842 observations. Table 1 shows the observed market shares of Gillette and Wilkinson Sword system blades and disposable razors. Out of the 4842 observations, 1973 observations correspond to system blade purchases and Table 2 shows the observed market shares conditional on buying either double or triple blade cartridges from either Gillette or Wilkinson Sword.

The covariates used for each household are indicators for age of the main shopper being 31-40 and 41-50, indicators for the main shopper's marital and employment status, and an indicator specifying the presence of more than one female in the household. Table 3 provides descriptive statistics. Further details are provided in Appendix B.

[9]Razor blades can be purchased in three forms: as disposable razors, as reusable razors sold with razor blade cartridges, or as razor blade cartridges for use with a previously purchased handle.

Razor Blade Type	Market Share
Gillette Blades	29.82%
Wilkinson Sword Blades	10.93%
Disposable Razors	59.25%

Table 1: 2004–2005 market shares.

Trading Company	Blade Type		Total
	Double Blade	Triple Blade	
Gillette	17.74%	55.45%	73.19%
Wilkinson Sword	9.22%	17.59%	26.81%
Total	26.96%	73.04%	100.00%

Table 2: Market shares conditional on purchasing double or triple blade cartridges from Gillette or Wilkinson Sword in 2004-2005.

For each observation in the sample the brand (b)-quality (y) combination of cartridges purchased is observed, in addition to the individual characteristics. For each purchase we observe the amount spent (w), the quantity purchased (q) and the specific pack size (v), as well as the month (m) and store (s) of purchase. For the observed purchases, the average unit price per cartridge p_{bymsv}, defined as the total amount spent divided by the total quantity purchased in a specific month and store, for each specific brand-quality combination and each pack size was computed as

$$p_{bymsv} = \frac{\sum_{i=1}^{N^{bymsv}} w_{bymsv,i}}{\sum_{i=1}^{N^{bymsv}} q_{bymsv,i}}. \tag{6.1}$$

where $i = 1, \ldots, N^{bymsv}$ and N^{bymsv} is the number of observations for each $bymsv$ combination. Regarding the unit prices faced by the consumers of the cartridges not bought, since these are not observed, we estimated them using the best linear predictor (BLP) under two

Age group		Marital status		Employment status		Number of females	
18-30	34.30%	Married	61.83%	Employed	70.76%	One female	43.04%
31-40	40.75%	Other	38.17%	Unemployed	29.24%	More than one	56.96%
41-50	24.95%						

Table 3: Main shopper characteristics for 2004-2005.

different specifications,

$$p_{bymsv} = \beta_0 + 1[b=2]\beta_1 + 1[y=2]\beta_2 + \sum_{\tilde{m}=2}^{24} 1[m=\tilde{m}]\kappa_m + \sum_{\tilde{s}=2}^{15} 1[s=\tilde{s}]\tau_s + \sum_{\tilde{v}=2}^{3} 1[v=\tilde{v}]\mu_v + \varepsilon, \quad (6.2)$$

$$p_{bymsv} = \beta_0 + 1[b=2]\beta_1 + 1[y=2]\beta_2 + \sum_{\tilde{m}=2}^{24} 1[m=\tilde{m}]\kappa_m + \sum_{\tilde{s}=2}^{15} 1[s=\tilde{s}]\tau_s + \varepsilon, \quad (6.3)$$

where ε denotes the residual of the BLP regression, $b = 2$ corresponds to Wilkinson Sword system blades, and $y = 2$ corresponds to triple blade cartridges. The intercept, β_0, corresponds to the price of a Gillette double blade cartridge sold in January 2004 in the store group category "All other" and for specification (6.2) in the small pack size. All of the other Greek letters denote BLP coefficients. These were estimated using all 1973 observations of cartridge purchases in our data for the 24 months spanning $2004 - 2005$, 15 stores, and three different pack sizes. To impute counterfactual prices for brand-quality combinations of system blades not purchased, the best linear predictors were matched to each observation according to the actual month, store and (in the case of specification (6.2)) pack size purchased.[10] The difference in specification (6.2) and specification (6.3) is that in the former we also control for pack size. We chose to differentiate between the two specifications as not all the blade-types and/or brands are observed offering all pack sizes. For example, double blade razors were only offered in five cartridge packs. In order to deal with this, pack size was categorized according to small, medium and large. See Appendix B for further details and descriptive statistics.

Using this data our application employs the partially ordered probit specification with $\bar{b} = 2$ and $\bar{y}_b = 2$ for each b considered in Section 5 with

$$U_{by} \equiv y \times (X\beta_b + V_b) - \delta_b - \gamma_b P_{by} \quad \text{if } (b,y) \neq (0,0), \qquad U_{00} \equiv 0,$$

where X are the individual specific indicators for the household's main shopper's age categories, marital status, employment, and the presence of multiple females in the household as previously described, and which here do not vary across brand.[11] Prices P_{by} on the other hand vary at both the product and individual level. The index $X\beta_b + V_b$ captures the

[10]The imputed prices are treated as observed variables; sample variation from the BLP regressions by which they were obtained is not taken into account in the subsequent analysis.

[11]It is straightforward to use specifications with a common price coefficient $\gamma = \gamma_1 = \gamma_2$. See Section 6.3 for further discussion in the context of the present application.

marginal effect on utility from an increase in quality Y when choosing brand b. All else equal, a higher value of this index increases both the utility of choosing the higher quality option ($Y = 2$) relative to the lower quality option ($Y = 1$) of brand b. It also increases the relative utility of brand b's offerings relative to those of its rival. The distribution of $V \equiv (V_1, V_2)$ is multivariate normal with parameters as indicated in Proposition 2. Thus, a positive value of the correlation parameter ρ indicates positive correlation in the taste for higher quality cartridges across the different brands, as would be expected if individuals who have higher unobserved taste for Gillette's triple blade cartridges also have higher unobserved taste for Wilkinson Sword's triple blade cartridges.

6.2 Empirical Implementation

Estimation and inference were carried out by maximum likelihood. Specifically, we report point estimates obtained by maximizing the likelihood

$$L_n(\zeta) \equiv \sum_{i=1}^{n} \left[\ln \wp_{b_i y_i}(z_i; \zeta)\right] \tag{6.4}$$

with respect to ζ, and 95% confidence intervals for each parameter component ζ_k obtained using both Wald and QLR statistics. Wald confidence intervals are computed as those values that come within 1.96 estimated standard deviations of the point estimate, where the Hessian form of the information matrix was used to estimate the asymptotic variance. The QLR confidence intervals are more time consuming to compute, but can be robust to a possible lack of point identification under certain conditions as shown by Chen, Christensen, and Tamer (2018), which are not verified here.[12] With point identification, both approaches are valid under standard conditions. In Appendix D we report the results of several Monte Carlo experiments comparing the two approaches.

[12]See Chen, Christensen, and Tamer (2018) for sufficient conditions for asymptotic validity of profile QLR confidence intervals for the *identified set* of parameter components ζ_k, each k, when ζ_k is not point identified. Note that a confidence interval that guarantees coverage of the entire identified set for ζ_k with pre-specified probability asymptotically is also guaranteed to contain the parameter ζ_k itself with at least the same probability asymptotically. This ensures that the confidence sets are asymptotically valid for ζ_k, although for coverage of the true parameter value only, it may in principle be possible to establish either weaker sufficient conditions or strictly smaller confidence intervals. Furthermore, while Chen, Christensen, and Tamer (2018) provide sufficient conditions for profile QLR confidence intervals to achieve at least the desired asymptotic coverage for identified sets of univariate *components* of the structural parameter, one could consider using the likelihood ratio statistic for inference on other objects, such as partial effects or various counterfactuals, although we are unaware of sufficient conditions for this to achieve valid inference under partial identification. These are questions we leave open to future research.

The steps taken to compute the QLR confidence intervals $\mathcal{C}^{LR}_{\alpha,k}$ were as follows, for each $k = 1, ..., \dim(\zeta)$ with $\alpha = 0.95$. First compute the unconstrained maximum of the log-likelihood denoted L^*_n and the quasi-likelihood ratio statistic

$$Q_{k,n}(\mu) \equiv \inf_{\zeta \in \Upsilon : \zeta_k = \mu} 2n \left[L^*_n - L_n(\zeta) \right], \tag{6.5}$$

was computed at values of μ over a grid of values $\mathcal{M}$.[13] The values

$$\begin{aligned} \overline{\mu_0} &\equiv \min \left\{ \mu \in \mathcal{M} : Q_{k,n}(\mu) \le \chi^2_{1,\alpha} \right\}, \quad \underline{\mu_0} \equiv \max \left\{ \mu \in \mathcal{M} : \mu < \overline{\mu_0} \right\}, \\ \underline{\mu_1} &\equiv \max \left\{ \mu \in \mathcal{M} : Q_{k,n}(\mu) \le \chi^2_{1,\alpha} \right\}, \quad \overline{\mu_1} \equiv \min \left\{ \mu \in \mathcal{M} : \mu > \underline{\mu_1} \right\}, \end{aligned}$$

were recorded, with $\chi^2_{1,\alpha}$ the α quantile of the chi-square distribution with one degree of freedom. Here $\overline{\mu_0}$ and $\underline{\mu_1}$ are the lowest and greatest values of μ on the grid $\mathcal{M}$ that pass the criterion $Q_{k,n}(\mu) \le \chi^2_{1,\alpha}$ required for $\mu \in \mathcal{C}^{LR}_{\alpha,k}$. The value $\underline{\mu_0}$ is the next lowest value to $\overline{\mu_0}$ on $\mathcal{M}$ while $\overline{\mu_1}$ is the next highest value to $\underline{\mu_1}$ on the grid. Then a minimal tolerance $\varepsilon > 0$ was set for the desired precision within which to compute each endpoint of $\mathcal{C}^{LR}_{\alpha,k}$ and the following steps were iterated.

1. Set $\widetilde{\mu} \equiv \left(\underline{\mu_0} + \overline{\mu_0} \right) / 2$ the halfway point between $\underline{\mu_0}$ and $\overline{\mu_0}$. Compute $Q_{k,n}(\widetilde{\mu})$.
2. If $Q_{k,n}(\widetilde{\mu}) \le \chi^2_{1,\alpha}$ then set $\overline{\mu_0} \equiv \widetilde{\mu}$. Otherwise set $\underline{\mu_0} \equiv \widetilde{\mu}$.
3. If $\left| \overline{\mu_0} - \underline{\mu_0} \right| > \varepsilon$ then return to step 1 and continue. Otherwise set the terminal value $\mu_0 \equiv \underline{\mu_0}$ and stop iterating.

Then the same steps were carried out for the upper bound of $\mathcal{C}^{LR}_{\alpha,k}$ by setting $\widetilde{\mu} \equiv \left(\underline{\mu_1} + \overline{\mu_1} \right) / 2$ and replacing $\overline{\mu_0}$ with $\underline{\mu_1}$ and $\underline{\mu_0}$ with $\overline{\mu_1}$ in the subsequent step. Here we let the terminal value be denoted μ_1. When the procedure is done, μ_0 and μ_1 serve as lower and upper bounds for $\mathcal{C}^{LR}_{\alpha,k}$.

6.3 Empirical Results

Tables 4 and 5 report maximum likelihood point estimates and confidence intervals constructed as described above using specifications (6.2) and (6.3), respectively, for counterfactual prices. The results in Table 4 lead to several observations. The coefficients γ_1 and γ_2

[13] The likelihood $L_n(\zeta)$ is as defined in (6.4) with $\wp_{by}(z_i; \zeta)$ defined in (5.3).

on price are positive, so that utility is measured to be decreasing in price. The coefficient γ_1 on price for Gillette is considerably smaller than the coefficient γ_2 for Wilkinson Sword, even after scaling by the estimate of σ_2.[14] The coefficient on the dummy variables for both age groups 31-40 and 41-50 are negative, as are their associated confidence intervals, with the exception of the coefficient on the 31-40 age group for Wilkinson Sword. This indicates a lower utility of system blade purchases of these age groups relative to the 18-30 age group. Likewise, the estimated coefficient on the more females indicator for both brands is negative. Coefficient estimates for employment and married dummy variables are negative and statistically indistinguishable from zero for both brands.

The estimated correlation coefficient between brand-specific unobservables is effectively one, indicating perfect correlation in unobserved preference for quality as reflected by blades per cartridge across the two different brands.[15] This indicates that consumers who have higher unobservable taste for the high quality product of one brand have higher unobservable taste for the higher quality product of the other brand too. This reflects choice patterns delivered by our fitted conditional choice probabilities with respect to variations in price, and is consistent with a setting in which consumers purchasing the higher quality offering of one brand would likely choose the higher quality offering of the rival brand if their chosen brand's product was not available or was prohibitively high-priced. While a high correlation in unobserved taste in the quality dimension thus seems reasonable in this application, a value of $\hat{\rho} \approx 1$ indicates roughly perfect correlation in unobserved taste for quality across brands. It also suggests that the population parameter ρ may be on the boundary of the parameter space, which can be problematic for inference. Indeed, the Hessian of the log-likelihood computed at the maximizing parameter vector was singular. Consequently, Wald confidence intervals are not reported for this specification. The QLR statistic can still be computed and the confidence intervals reported here use the χ^2_1 critical value which is valid if the parameter is on the interior of the parameter space.[16] The QLR confidence interval

[14] As noted previously, it is straightforward to impose that $\gamma = \gamma_1 = \gamma_2$. This was tried in the present application, but produced estimates implying essentially zero variance in V_2. Thus, results are reported using the more flexible specification in which γ_1 and γ_2 may differ, which appears to better fit the data. This may be down to several factors, such as unobservable factors in brand–quality offerings or heterogeneity in consumer consideration sets (see e.g. Barseghyan, Coughlin, Molinari, and Teitelbaum (2021)) possibly due to the need to have previously purchased a compatible handle, or lack of availability of certain combinations of products in all stores. Such issues are further discussed at the end of this section. A potentially worthwhile generalization, not attempted here, could also be to specify γ as a brand-constant random coefficient that varies across consumers.

[15] The maximum likelihood estimate was 0.99999992, effectively indistinguishable from 1.

[16] If in fact the population value of ρ is one, alternative critical values are needed, see for example Self and

obtained for ρ is very tightly concentrated around 1.

Table 5 reports results obtained using specification (6.3) for counterfactual prices. The estimate of the correlation coefficient ρ between brand-specific unobservables is again very close to one, again indicating near perfect correlation in preference for quality (blades per cartridge) across the two different brands. However, the likelihood-maximizing value of ρ was slightly lower than was found using specification (6.2) for counterfactual prices. The Hessian was nonsingular and both Wald and QLR confidence intervals for each parameter are reported.[17] The point estimate for γ_1, the price coefficient for Gillette, is negative, indicating that utility is increasing in price, although its magnitude is small and the confidence intervals indicate it is nearly statistically indistinguishable from zero. The coefficient estimate on price for Wilkinson Sword is positive, and statistically significantly different from zero, indicating that utility from purchasing these products is decreasing in price. For the most part, the signs of coefficient estimates and confidence intervals for other variables accord qualitatively with those of the prior specification. Two slight exceptions are that although $\beta_{2Age31-40}$ and $\beta_{2married}$ are again estimated to be negative, their associated confidence intervals now lie fully below zero. The estimate of σ_2 is slightly smaller than it was using the previous specification, but of similar magnitude.

Table 6 reports features of the estimated own- and cross-price elasticities

$$\eta_{bykl}(z_i;\zeta) \equiv \frac{\partial \wp_{by}(z_i;\zeta)}{\partial p_{k\ell}} \cdot \frac{p_{k\ell}}{\wp_{by}(z_i;\zeta)} \tag{6.6}$$

for each (b, y) product combination and for the outside good implied by the parameter estimates reported in Table 4. For each household i the corresponding elasticity of the choice probability with respect to each price was computed.[18] Table 6 displays the resulting means and 0.2, 0.5, and 0.8 quantiles of these elasticities.[19] Relative to the means, the quantiles illustrate considerable heterogeneity in household substitution patterns. Focusing first on

Liang (1987) and Andrews (1999).

[17]Small perturbations of ρ near the maximizing parameter vector were investigated and confirmed to result in a decrease in the log-likelihood.

[18]Note that elasticities were computed conditional on individual household covariates z_i. They are not elasticities of an aggregate demand function. Mean estimates are sensitive to the presence of z_i with very small estimated conditional choice probabilities $\wp_{by}(z_i;\widehat{\zeta})$.

[19]Fitted shares less than 10^{-8} were rounded to zero, and corresponding elasticity estimates were likewise set to zero to avoid approximation error. Code for computing these elasticities is included with the code used to compute the likelihood and produce the Monte Carlo results in Appendix D, which can be downloaded at `https://sites.google.com/site/amr331/home/por-code`. An additional link on this page provides a pdf web supplement that derives the expressions that were used to compute the elasticities.

Parameter	ML Point Estimate	QLR 95% CI
γ_1	0.0613	[-0.0380, 0.1736]
γ_2	4.2811	[3.4006, 5.8717]
δ_1	-0.2019	[-0.2485, -0.1611]
δ_2	-4.0530	[-5.3888, -3.3035]
$\beta_{1Age31-40}$	-0.1567	[-0.2359, -0.0772]
$\beta_{1Age41-50}$	-0.2261	[-0.3208, -0.1323]
$\beta_{1Married}$	-0.0445	[-0.1092, 0.0202]
$\beta_{1Employed}$	-0.0381	[-0.1072, 0.0320]
$\beta_{1Females}$	-0.2238	[-0.2920, -0.1560]
$\beta_{2Age31-40}$	-0.2394	[-0.4837, 0.0042]
$\beta_{2Age41-50}$	-0.3626	[-0.6418, -0.0869]
$\beta_{2Married}$	-0.1859	[-0.3859, 0.0128]
$\beta_{2Employed}$	-0.1138	[-0.3316, 0.0915]
$\beta_{2Females}$	-0.6102	[-0.8768, -0.4079]
ρ	1.0000	[0.9994, 1.0000]
σ_2	2.6938	[2.3475, 3.3302]

Table 4: 95% confidence intervals with prices as specified in (6.2).

Parameter	ML Point Estimate	Wald 95% CI	QLR 95% CI
γ_1	-0.0643	[-0.1250, -0.0037]	[-0.1157, -0.0037]
γ_2	2.6258	[2.3509, 2.9006]	[2.3448, 2.9130]
δ_1	-0.1650	[-0.1855, -0.1445]	[-0.1850, -0.1467]
δ_2	-2.5891	[-2.7745, -2.4037]	[-2.7805, -2.3872]
$\beta_{1Age31-40}$	-0.1595	[-0.2381, -0.0810]	[-0.2369, -0.0822]
$\beta_{1Age41-50}$	-0.2163	[-0.3107, -0.1218]	[-0.3087, -0.1240]
$\beta_{1Married}$	-0.0814	[-0.1439, -0.0190]	[-0.1435, -0.0192]
$\beta_{1Employed}$	-0.0585	[-0.1323, 0.0154]	[-0.1266, 0.0098]
$\beta_{1Females}$	-0.2378	[-0.3078, -0.1678]	[-0.3045,-0.1710]
$\beta_{2Age31-40}$	-0.3057	[-0.4811, -0.1303]	[-0.4763, -0.1361]
$\beta_{2Age41-50}$	-0.4181	[-0.6270, -0.2092]	[-0.6209, -0.2171]
$\beta_{2Married}$	-0.1937	[-0.3360, -0.0514]	[-0.3319 ,-0.0553]
$\beta_{2Employed}$	-0.1127	[-0.2741, 0.0488]	[-0.2620 ,0.0368]
$\beta_{2Females}$	-0.5249	[-0.6803, -0.3694]	[-0.6720, -0.3800]
ρ	0.9998	[0.9996, 1.0000]	[0.9996, 1.0000]
σ_2	2.1580	[2.0518, 2.2641]	[2.0334, 2.2734]

Table 5: 95% confidence intervals with prices as specified in (6.3).

mean own-price elasticities, sales of Gillette cartridges ($B = 1$) appear to be less sensitive to changes in own-price than Wilkinson Sword cartridges ($B = 2$). However, shifting focus to the quantiles of elasticities shows a more nuanced picture. For example, the estimated mean own-price elasticity for alternative $(2, 1)$ is roughly -65, which is extremely high, but the estimated median elasticity is zero and the estimated 0.2 quantile is -2.136. In addition, the 0.1 and 0.05 quantiles (not reported in the table) are roughly -32 and -67, respectively. The estimates suggest that the majority of households are not considering this product at the margin. For these households $\wp_{21}(z_i;\zeta)$ is unaffected by a local change in price. However there is a small subset of households that in fact are quite sensitive to a change in the price of this product. Indeed this is consistent with this product offering having by far the smallest market share in our data. A small price increase that leads to even a handful of households no longer purchasing can produce a large percentage change in the share purchasing. That said, elasticities featuring extremely small conditional choice probabilities $\wp_{by}(z_i;\zeta)$ in the denominator in (6.6) may be especially sensitive to sampling variation and should thus be interpreted with caution.

A qualitatively similar observation holds for η_{2122}; the purchase probability $\wp_{21}$ is insensitive to a change in p_{22} for households not locally considering product $(2,1)$, but purchase probabilities for the segment of households that are purchasing or are locally close to purchasing this product respond considerably to changes in p_{22}. For these households, product $(2,2)$ appears to be the closest substitute. The large estimated mean own-price elasticity of -3.887 for $\wp_{22}$ presents a more uniform level of sensitivity to price changes, as the quantiles are much less disperse. This product has nearly double the market share of product $(2,1)$ and is generally the most expensive product available.

Cross-price elasticity estimates reflect various additional substitution patterns. For instance, the elasticity estimates for Gillette's double blade offering $(1,1)$ with respect to Wilkinson blade cartridges ($B=2$) are zero for most consumers, up to at least the 0.8 quantile. However, the substantially higher mean estimate with respect to p_{21} illustrates that there is a segment of consumers who are locally sensitive to price changes in Wilkinson's double blade razors; in this sense product $(2,1)$ is a closer substitute for $(1,1)$ than is product $(2,2)$. On the other hand, for Gillette's triple blade razors, product $(1,2)$, there are more consumers sensitive to price changes in Wilkinson's triple blade cartridges $(2,2)$ than Wilkinson's double blade cartridges $(2,1)$. Given the ordered nature of the product offerings, the probability of purchasing the outside good, disposable razors, is somewhat sensitive to changes in the price of double blade razors ($Y=1$), but is insensitive to changes in the price of triple blade razors ($Y=2$).

Before concluding, it should be noted that these observations come with at least two important caveats. First of all, our model is static. We do not model consumers' purchase or prior possession of handle for any brand/quality offering. Ownership of a handle for product

(b, y) that is not compatible with other products may be an important factor in consumer demand and observed substitution patterns. Addressing this would require a dynamic model and data on handle purchases that may have occurred well in the past, which we do not have. Second, we do not actually observe the menu of products each consumer faces when making their purchase. Even if we were able to accurately impute counterfactual prices based on prices paid by other consumers in a given store and month, it is possible that products a consumer did not purchase were actually not available in the store at the time the consumer was shopping. This could be problematic if some stores regularly did not carry or sold out of certain products, rendering imputed prices invalid.

Despite these caveats, the application illustrates the ability of the partially ordered response model to produce consumer substitution patterns that accord with the dual horizontal and vertical dimensions of product differentiation. To illustrate this point, for the sake of comparison we report in Table 15 in Appendix C elasticities instead obtained using the alternative specific conditional logit model of McFadden (1974), implemented as the asclogit command in Stata, using price specification (6.2). Relative to our model, this model ignores the vertical dimension of differentiation, thus treating all products as horizontally differentiated. In comparison to the average elasticities in Table 6, the magnitude of the average own-price elasticities for Gillette products ($B = 1$) in Table 15 is larger, indicating a greater average substitution effect following an increase in own price, while the reverse is true for Wilkson products ($B = 2$). Comparing the quantiles the picture is also quite different. An implication of the logit model is that no matter what the value of z_i for a household, there is a strictly positive probability that a marginal change in the price of the product it purchases will induce it to switch to any one of the other alternatives. Consequently, at each quantile of price elasticities, there is a non-zero proportion of households considering each product. An even more striking difference is with respect to the cross-price elasticities. While in our model the cross-price elasticities are different, the cross-price elasticities for product (b, y) with respect to $p_{k\ell}$ implied by the alternative specific conditional logit model are the same regardless of (b, y). For example, the mean cross-price elasticity of product (b, y) induced by a change in the price of the Gillette double blade cartridge (product $(1, 1)$) reported in Table 15 is estimated as 0.0874, across *all* $(b, y) \neq (1, 1)$. This is restrictive, as we would expect individuals to respond to price changes of a given product (e.g. $(1, 1)$) differently depending on which product they are currently purchasing.[20] Our model allows this, as the

[20]This issue as well as other implications of the logit elasticities have been previously discussed in the literature, for example in Nevo (2000).

first column of Table 6 indicates that the share of $(1, 2)$ and the outside alternative are both locally sensitive to changes in p_{11}, although the corresponding cross-price elasticities differ. Furthermore, Table 6 indicates consumers choosing the shares of $(2, 1)$ and $(2, 2)$ are locally insensitive to changes in p_{11}. More generally it is easy to see from the columns of Table 6 that cross-price elasticities reported with respect to any given product price may vary.

The estimated elasticities further indicate interesting substitution patterns within and across brands. Looking at the rows for $\wp_{11}$ and $\wp_{12}$, we generally see non-zero cross-price elasticities for consumers choosing Gillette's alternative product offering, but zero quantiles of many of the cross-price elasticities with respect to Wilkinson's products. This suggests most consumers who purchase Gillette cartridges are more responsive to local changes in the price of Gillette's other product than they are to local changes in Wilkinson's prices. Considering the rows for $\wp_{21}$ and $\wp_{22}$ we instead see that consumers purchasing Wilkinson's products are comparatively insensitive to changes in the prices of other products, with exceptions at some quantiles of estimated cross-price elasticities.

7 Conclusion

In this paper we proposed a new discrete choice model for partially ordered alternatives, applicable when discrete choices are differentiated along both vertical and horizontal dimensions. We provided general characterizations of the identified set of structures admitted by the model and conditional choice probabilities under mild shape restrictions that induce an ordered (i.e. vertically differentiated) choice problem holding the horizontal dimension of choice fixed. We considered specialized results for a partially ordered probit model in which brand-specific unobservables are restricted to be multivariate normal with parameterized covariance matrix. Closed form expressions for choice probabilities were obtained in a duopoly (two-brand) setting, useful for computing the log-likelihood. An empirical illustration was provided using data on razor blade cartridge purchases in a setting that features two dominant competing firms with vertically differentiated products. The application demonstrated the model's ability to capture interesting substitution patterns commensurate with the two levels of differentiation, and which reflect underlying heterogeneity in preferences.

Estimating such patterns as accurately as possible can be important for measuring welfare, predicting reactions to the introduction of new products, and for modeling firm competition. Indeed, future analysis combining demand modeled by way of a vertically and horizontally ordered choice model with a model of equilibrium firm behavior could be useful

	Price Elasticities											
$\frac{\partial \log \wp_{by}}{\partial \log p_{k\ell}}$		p_{11}			p_{12}			p_{21}			p_{22}	
$\wp_{11}$												
mean		-0.528			0.398			0.902			0.001	
quantiles	-0.596	-0.543	-0.422	0.362	0.421	0.455	0.000	0.000	0.000	0.000	0.000	0.000
$\wp_{12}$												
mean		0.187			-0.658			20.549			1.142	
quantiles	0.061	0.072	0.080	-0.152	-0.134	-0.116	0.000	0.000	0.000	0.000	1.332	1.629
$\wp_{21}$												
mean		0.000			0.801			-64.979			19.384	
quantiles	0.000	0.000	0.000	0.000	0.000	0.000	-2.136	0.000	0.000	0.000	0.000	1.003
$\wp_{22}$												
mean		0.000			0.033			0.564			-3.887	
quantiles	0.000	0.000	0.000	0.000	0.040	0.045	0.000	0.000	2.245	-4.243	-3.848	-3.446
$\wp_0$												
mean		0.033			0.000			0.025			0.000	
quantiles	0.028	0.033	0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6: Estimated means and 0.2, 0.5, and 0.8 quantiles of household elasticities.

for estimating such quantities. In this way, the choice model developed here could potentially be applied to study equilibrium pricing by multiproduct firms engaging in second degree price discrimination. This could in turn be used to empirically measure the direction and magnitude of welfare effects of competition in markets with non-linear pricing, for which the incorporation of both horizontal and vertical differentiation is important.

A Proofs

Proof of Proposition 1. First consider any $(u_0, G_0) \in \mathcal{S}_0$. By the same argument as when there is point identification we have for almost every $z \in \mathcal{Z}$,

$$E\left[\ln G_0\left(\mathcal{V}_{BY}\left(z; u_0\right)\right)|Z = z\right] \geq E\left[\ln G\left(\mathcal{V}_{BY}\left(z; u\right)\right)|Z = z\right] \tag{A.1}$$

for all $(u, G) \in \mathcal{S}$. Thus $\mathcal{S}_0$ is contained in the set of maximizers of $Q(u, G)$. Consider now $\left(\tilde{u}, \tilde{G}\right) \notin \mathcal{S}_0$. Then for some $(b, y) \in \mathcal{M}_{BY}$ there exists a positive measure set $\mathcal{Z}^* \left(\tilde{u}, \tilde{G}\right)$ as defined in (3.7) on which $\tilde{G}\left(\mathcal{V}_{by}\left(z; \tilde{u}\right)\right) \neq G_0\left(\mathcal{V}_{BY}\left(z; u_0\right)\right) = f_z^0 (b, y)$ for at least one (b, y) pair. We therefore have

$$\forall z \in \mathcal{Z}^* \left(\tilde{u}, \tilde{G}\right), \, E\left[\ln G_0\left(\mathcal{V}_{BY}\left(z; u_0\right)\right) | Z = z\right] > E\left[\ln \tilde{G}\left(\mathcal{V}_{BY}\left(z; \tilde{u}\right)\right) | Z = z\right].$$

Combining this with (A.1) it follows that $Q(u_0, G_0) > Q\left(\tilde{u}, \tilde{G}\right)$, completing the proof. ■

Proof of Theorem 1. From the utility maximization hypothesis, (b, y) is chosen if and only if it maximizes $u(b, y, z, v_b)$. This is so if and only (i) (b, y) provides higher utility than that delivered by all within brand options $\{u(b, \tilde{y}, z, v_b) : \tilde{y} \neq y\}$, and (ii) (b, y) provides higher utility than that delivered by all alternative brand options $\left\{u\left(\tilde{b}, \tilde{y}, z, v_{\tilde{b}}\right) : \left(\tilde{b}, \tilde{y}\right) \neq (b, y)\right\}$.

Condition (i) requires that y maximizes $u(b, \cdot, z, v_b)$ for the stated brand b, that is $Y_b^* = y$. Given the single-crossing property of Restriction A6(iii) we can apply Theorem 4 of Milgrom and Shannon (1994), implying that Y_b^* is nondecreasing in v_b. It follows that for each $y \in \{0, ..., \bar{y}_b + 1\}$, there is a nondecreasing sequence of thresholds $\{g_b(y) : y = (0, ..., \bar{y}_b + 1)\}$ such that $Y_b^* = y$ if and only if $v_b \in (g_b(y), g_b(y+1)]$, where possibly $g_b(y) = g_b(y+1)$ if alternative (b, y) is never chosen. That $g_b(0) \equiv -\infty$ and $g_b(\bar{y}_b + 1) \equiv \infty$ follows from $y = 0$ and $\bar{y}_b$ being the lowest and highest feasible values of y.

Condition (ii) stems from Restriction A6(ii), strict monotonicity of $u(b, y, z, v_b)$ in v_b for each b. The consumer will choose brand b if and only if for any other brand d, the utility from choosing (b, Y_b^*) exceeds that from choosing (d, Y_d^*), that is if

$$u(b, Y_b^*, z, v_b) > \max_{y \in \mathcal{Y}_d} u(d, y, z, v_d), \text{ if } b < d, \tag{A.2}$$

$$u(b, Y_b^*, z, v_b) \geq \max_{y \in \mathcal{Y}_d} u(d, y, z, v_d), \text{ if } b > d. \tag{A.3}$$

By A6(ii) it follows that

$$u_d^*(z, v_d) \equiv \max_{y \in \mathcal{Y}_d} u(d, y, z, v_d)$$

is strictly monotone and thus invertible in v_d. Inequalities (A.2) and (A.3) thus simplify to

$$q_d\{u(b, Y_b^*, z, v_b); z\} > v_d, \text{ if } b < d,$$

$$q_d\{u(b, Y_b^*, z, v_b); z\} \geq v_d, \text{ if } b > d,$$

where $q_d(\cdot; z)$ denotes the inverse of $u_d^*(z, v_d)$ with respect to v_d, i.e. for any (z, v_d),

$$q_d\left(u_d^*(z, v_d); z\right) = v_d.$$

Then we have the inequalities (3.3) and (3.4) with

$$h_{b,d}(y) \equiv q_d\left\{u\left(b, Y_b^*, z, v_b\right); z\right\},$$

for each pair $b \neq d$. The integral (3.5) for the conditional choice probabilities then follows immediately from their definition $\wp_{by}(z; \mathcal{S}) \equiv G\left(\mathcal{V}_{by}(z; u)\right)$. ■

Proof of Theorem 2. It is straightforward to verify that the function

$$h(v, \theta) \equiv f_V(v; \Sigma)\, 1\left[v \in \mathcal{V}_{by}(z; \theta)\right]$$

is log-concave in (v, θ). This follows from log-concavity of $f_V(v; \Sigma)$ in v and log-concavity of $1\left[v \in \mathcal{V}_{by}(z; \theta)\right]$ in (v, θ), which is easy to establish given $\mathcal{V}_{by}(z; \theta)$ comprises a system of linear inequalities in (v, θ). By Theorem 6 of Prekopa (1973) it then follows that

$$\int h(v, \theta)\, dv$$

is log-concave in θ and concavity of $\mathcal{L}(\theta, G)$ follows. ■

Proof of Theorem 3. Let $\tilde{\theta} \neq \theta$ and for each $b \in \mathcal{B}$ let $\theta_b \equiv \left(\delta_b, \gamma_b, \beta_b'\right)'$ and $\tilde{\theta}_b \equiv \left(\tilde{\delta}_b, \tilde{\gamma}_b, \tilde{\beta}_b'\right)'$, $\theta_{-b} \equiv \left(\theta_1', ..., \theta_{b-1}', \theta_{b+1}', ..., \theta_{\bar{b}}'\right)'$, and likewise for $\tilde{\theta}_{-b}$. Let k_b denote the number of components of θ_b.

Let

$$\mathbf{z} \equiv \begin{pmatrix} \tilde{z}_{1,1} & \cdots & \tilde{z}_{1,k_1} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \tilde{z}_{2,1} & \cdots & \tilde{z}_{2,k_2} & 0 & \cdots & 0 & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \tilde{z}_{\bar{b},1} & \cdots & \tilde{z}_{\bar{b},k_{\bar{b}}} \end{pmatrix}$$

and let $\mathbf{z}_{-b}$ denote the matrix $\mathbf{z}$ with the b_{th} row removed. Let $\mathbf{z}_b$ denote the b_{th} row of $\mathbf{z}$ excluding zero entries. As in the main text, z (not in bold font) is used to denote a realization of all exogenous variables excluding repetitions.

Define the sets

$$S_b^+ \equiv \left\{z \in \mathcal{Z}^* : \mathbf{z}_b\left(\tilde{\theta}_b - \theta_b\right) > 0 \text{ and } \mathbf{z}_{-b}\left(\tilde{\theta}_{-b} - \theta_{-b}\right) \geq 0\right\},$$

$$S_b^- \equiv \left\{ z \in \mathcal{Z}^* : \mathbf{z}_b \left(\tilde{\theta}_b - \theta_b \right) < 0 \text{ and } \mathbf{z}_{-b} \left(\tilde{\theta}_{-b} - \theta_{-b} \right) \leq 0 \right\}.$$

Let $\Phi_{\bar{b}} (\mathbf{z}\theta; \Sigma)$ denote the cumulative distribution function for the $\bar{b}$-variate distribution $G(\cdot, \Sigma)$ evaluated at $\mathbf{z}\theta$. For any $z \in S_b^+$ we have that

$$\Phi_{\bar{b}} \left(\mathbf{z}\tilde{\theta}; \Sigma \right) > \Phi_{\bar{b}} (\mathbf{z}\theta; \Sigma) = f_z^0 (0, 0), \tag{A.4}$$

and likewise for any $z \in S_b^-$,

$$\Phi_{\bar{b}} \left(\mathbf{z}\tilde{\theta}; \Sigma \right) < \Phi_{\bar{b}} (\mathbf{z}\theta; \Sigma) = f_z^0 (0, 0), \tag{A.5}$$

where $f_z^0 (0, 0) = \mathbb{P} \{ (B, Y) = (0, 0) | Z = z \}$. The probability that $Z \in S_b \equiv S_b^+ \cup S_b^-$ is

$$\begin{aligned} \mathbb{P} \{ Z \in S_b \} &= \mathbb{P} \left\{ Z \in S_b^+ \right\} + \mathbb{P} \left\{ Z \in S_b^- \right\} \\ &= \left(\begin{array}{c} \mathbb{P} \left\{ \mathbf{Z}_{-b} \left(\tilde{\theta}_{-b} - \theta_{-b} \right) \geq 0 | \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) > 0, Z \in \mathcal{Z}^* \right\} \mathbb{P} \left\{ \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) > 0, Z \in \mathcal{Z}^* \right\} \\ + \mathbb{P} \left\{ \mathbf{Z}_{-b} \left(\tilde{\theta}_{-b} - \theta_{-b} \right) \leq 0 | \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) < 0, Z \in \mathcal{Z}^* \right\} \mathbb{P} \left\{ \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) < 0, Z \in \mathcal{Z}^* \right\} \end{array} \right) \end{aligned}$$

At least one of $\mathbb{P} \left\{ \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) > 0, Z \in \mathcal{Z}^* \right\}$ and $\mathbb{P} \left\{ \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) < 0, Z \in \mathcal{Z}^* \right\}$ are strictly positive because of condition (i), and $\mathbb{P} \{ Z \in \mathcal{Z}^* \} > 0$, and it follows that at least one of

$$\mathbb{P} \left\{ \mathbf{Z}_{-b} \left(\tilde{\theta}_{-b} - \theta_{-b} \right) \geq 0 | \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) > 0, Z \in \mathcal{Z}^* \right\} \mathbb{P} \left\{ \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) > 0, Z \in \mathcal{Z}^* \right\},$$

$$\mathbb{P} \left\{ \mathbf{Z}_{-b} \left(\tilde{\theta}_{-b} - \theta_{-b} \right) \leq 0 | \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) < 0, Z \in \mathcal{Z}^* \right\} \mathbb{P} \left\{ \mathbf{Z}_b \left(\tilde{\theta}_b - \theta_b \right) < 0, Z \in \mathcal{Z}^* \right\},$$

or both must be strictly positive by condition (ii). Therefore $\mathbb{P} \{ Z \in S_b \} > 0$, implying that $\tilde{\theta}$ is observationally distinct from θ since for each $z \in S_b$, $f_z^0 (0, 0) \neq \Phi_{\bar{b}} \left(\mathbf{z}\tilde{\theta}; \Sigma \right)$. ■

Before proving Proposition 2, the following Lemma is first proven.

Lemma 1 *When $\bar{y}_b = 2$ for each b, then (5.2) can be simplified to*

$$h_b (y, z, v, \theta) = 1 \left[v < z_{by}^* \right] m_{by}^- (z, v, \theta) + 1 \left[v \geq z_{by}^* \right] m_{by}^+ (z, v, \theta), \tag{A.6}$$

where

$$\begin{aligned} m_{by}^- (z, v, \theta) &\equiv y (x_b \beta_b + v) + \alpha_{d1} - \alpha_{by} - x_d \beta_d, & \text{(A.7)} \\ m_{by}^+ (z, v, \theta) &\equiv \frac{1}{2} [y (x_b \beta_b + v) + \alpha_{d2} - \alpha_{by}] - x_d \beta_d, & \text{(A.8)} \end{aligned}$$

Proof. Since $\bar{y}_b = 2$, (5.2) simplifies to

$$h_b(y,z,v,\theta) = \min_{\tilde{y}\in\{1,\ldots,\bar{y}_d\}} \frac{1}{\tilde{y}}\left[y\left(x_b\beta_b+v\right)-\left(\alpha_{by}-\alpha_{d\tilde{y}}\right)\right]-x_d\beta_d \tag{A.9}$$

$$= \min\left\{m^-_{by}(z,v,\theta), m^+_{by}(z,v,\theta)\right\}. \tag{A.10}$$

Both $m^-_{by}(z,v,\theta)$ and $m^+_{by}(z,v,\theta)$ are linear and strictly increasing in v. Setting $m^-_{by}(z,v,\theta) = m^+_{by}(z,v,\theta)$ reveals that the two functions are equal at

$$v = z^*_{by} \equiv \frac{\alpha_{d2}+\alpha_{by}-2\alpha_{d1}}{y} - x_b\beta_b,$$

and since $m^-_{by}(z,v,\theta)$ has a larger slope with respect to v, it follows that for all $v < z^*_{by}$, $m^-_{by}(z,v,\theta) < m^+_{by}(z,v,\theta)$, while for all $v > z^*_{by}$, $m^-_{by}(z,v,\theta) > m^+_{by}(z,v,\theta)$. Thus (A.10) simplifies to (A.6), completing the proof. ■

Proof of Proposition 2. The starting point is (5.1):

$$\wp_{by}(z,\zeta) = \frac{1}{\sigma_b}\int_{g_b(y;z,\theta)}^{g_b(y+1;z,\theta)} \Phi\left(\frac{h_b(y,z,v,\theta)-\rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv,$$

which is broken into three cases, depending on whether z^*_{by} lies below, inside, or above the interval $[g_b(y;z,\theta), g_b(y+1;z,\theta)]$ on which the integral is to be evaluated.

1. $g_b(y;z,\theta) < z^*_{by} < g_b(y+1;z,\theta)$.

$$\wp_{by}(z,\zeta) = \sigma_b^{-1}\int_{g_b(y;z,\theta)}^{z^*_{by}} \Phi\left(\frac{m^-_{by}(z,v,\theta)-\rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv + \sigma_b^{-1}\int_{z^*_{by}}^{g_b(y+1;z,\theta)} \Phi\left(\frac{m^+_{by}(z,v,\theta)-\rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv \tag{A.11}$$

2. $g_b(y;z,\theta) \le g_b(y+1;z,\theta) \le z^*_{by}$.

$$\wp_{by}(z,\zeta) = \sigma_b^{-1}\int_{g_b(y;z,\theta)}^{g_b(y+1;z,\theta)} \Phi\left(\frac{m^-_{by}(z,v,\theta)-\rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv \tag{A.12}$$

3. $z_{by}^{*} \leq g_b(y;z,\theta) \leq g_b(y+1;z,\theta)$.

$$\wp_{by}(z,\zeta) = \sigma_b^{-1} \int_{g_b(y;z,\theta)}^{g_b(y+1;z,\theta)} \Phi\left(\frac{m_{by}^{+}(z,v,\theta) - \rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv \tag{A.13}$$

The expressions in each case simplify as follows, using (A.7) and (A.8) and a change of variables substitution for $\frac{v}{\sigma_b}$.

$$\begin{aligned}\sigma_b^{-1} \int_{g_b(y;z,\theta)}^{z_{by}^{*}} &\Phi\left(\frac{m_{by}^{-}(z,v,\theta) - \rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv \\ &= \int_{\sigma_b^{-1}g_b(y;z,\theta)}^{\sigma_b^{-1}z_{by}^{*}} \Phi\left(\frac{yx_b\beta_b + \alpha_{d1} - \alpha_{by} - x_d\beta_d + (\sigma_b y - \rho\sigma_d)v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi(v)\,dv\end{aligned} \tag{A.14}$$

$$\begin{aligned}\sigma_b^{-1} \int_{z_{by}^{*}}^{g_b(y+1;z,\theta)} &\Phi\left(\frac{m_{by}^{+}(z,v,\theta) - \rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv \\ &= \int_{\sigma_b^{-1}z_{by}^{*}}^{\sigma_b^{-1}g_b(y+1;z,\theta)} \Phi\left(\frac{\frac{1}{2}\left[yx_b\beta_b + \alpha_{d2} - \alpha_{by}\right] - x_d\beta_d + \left(\frac{1}{2}\sigma_b y - \rho\sigma_d\right)v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi(v)\,dv.\end{aligned} \tag{A.15}$$

$$\begin{aligned}\sigma_b^{-1} \int_{g_b(y;z,\theta)}^{g_b(y+1;z,\theta)} &\Phi\left(\frac{m_{by}^{-}(z,v,\theta) - \rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv \\ &= \int_{\sigma_b^{-1}g_b(y;z,\theta)}^{\sigma_b^{-1}g_b(y+1;z,\theta)} \Phi\left(\frac{yx_b\beta_b + \alpha_{d1} - \alpha_{by} - x_d\beta_d + (\sigma_b y - \rho\sigma_d)v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi(v)\,dv\end{aligned} \tag{A.16}$$

$$\sigma_b^{-1} \int\limits_{g_b(y;z,\theta)}^{g_b(y+1;z,\theta)} \Phi\left(\frac{m_{by}^{+}(z,v,\theta) - \rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv$$

$$= \int\limits_{\sigma_b^{-1}g_b(y;z,\theta)}^{\sigma_b^{-1}g_b(y+1;z,\theta)} \Phi\left(\frac{\frac{1}{2}\left[yx\beta_b + \alpha_{d2} - \alpha_{by}\right] - x_d\beta_d + \left(\frac{1}{2}\sigma_b y - \rho\sigma_d\right)v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi(v)\,dv \quad \text{(A.17)}$$

Page 403 of Owen (1980) gives us formula 10,010.4:

$$\int\limits_{h}^{k} \Phi(c_1 + c_2 z)\,\phi(z)\,dz = \Lambda(k, c_1, c_2) - \Lambda(h, c_1, c_2) \quad \text{(A.18)}$$

where the function $\Lambda(\cdot,\cdot,\cdot)$ is given by

$$\Lambda(k, c_1, c_2) = \int\limits_{-\infty}^{\frac{c_1}{\sqrt{c_2^2+1}}} \phi(z)\,\Phi\left(k\sqrt{c_2^2+1} + c_2 z\right)dz. \quad \text{(A.19)}$$

Formula 10,010.1 on page 402 of Owen (1980) is

$$\int\limits_{-\infty}^{y} \phi(z)\,\Phi(a + bz)\,dz = \Phi_2\left(\frac{a}{\sqrt{1+b^2}}, y; \frac{-b}{\sqrt{1+b^2}}\right). \quad \text{(A.20)}$$

Applying this formula to (A.19) with

$$y = \frac{c_1}{\sqrt{c_2^2+1}}, \qquad a = k\sqrt{c_2^2+1}, \qquad b = c_2$$

gives

$$\Lambda(k, c_1, c_2) = \Phi_2\left(k, \frac{c_1}{\sqrt{c_2^2+1}}; \frac{-c_2}{\sqrt{1+c_2^2}}\right). \quad \text{(A.21)}$$

Define now

$$\begin{aligned} c_1^- &\equiv \frac{yx_b\beta_b + \alpha_{d1} - \alpha_{by} - x_d\beta_d}{\sigma_d\sqrt{1-\rho^2}}, \quad c_2^- \equiv \frac{\sigma_b y - \rho\sigma_d}{\sigma_d\sqrt{1-\rho^2}}, \\ c_1^+ &\equiv \frac{yx_b\beta_b + \alpha_{d2} - \alpha_{by} - 2x_d\beta_d}{2\sigma_d\sqrt{1-\rho^2}}, \quad c_2^+ \equiv \frac{\sigma_b y - 2\rho\sigma_d}{2\sigma_d\sqrt{1-\rho^2}}. \end{aligned}$$

as well as

$$\tilde{\Delta}(h,k,c_1,c_2) \equiv \Lambda(k,c_1,c_2) - \Lambda(h,c_1,c_2). \tag{A.22}$$

Referring back to (A.18), substitution of c_2 with those coefficients multiplying v and substitution of c_1 with those terms not multiplying v in the integrands on the rights hand side of (A.14) - (A.17) combined with (A.11) - (A.13) gives the following expression for conditional choice probabilities according to where z^*_{by} lies with respect to the interval $[g_b(y;z,\theta), g_b(y+1;z,\theta)]$.

1. $g_b(y;z,\theta) < z^*_{by} < g_b(y+1;z,\theta)$.

$$\begin{aligned}\wp_{by}(z,\zeta) &= \sigma_b^{-1}\left\{\begin{array}{l}\displaystyle\int_{g_b(y;z,\theta)}^{z^*_{by}} \Phi\left(\frac{m^-_{by}(z,v,\theta)-\rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv \\ \displaystyle+\int_{z^*_{by}}^{g_b(y+1;z,\theta)} \Phi\left(\frac{m^+_{by}(z,v,\theta)-\rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv\end{array}\right\} \\ &= \tilde{\Delta}\left(\sigma_b^{-1}g_b(y;z,\theta), \sigma_b^{-1}z^*_{by}, c_1^-, c_2^-\right) + \tilde{\Delta}\left(\sigma_b^{-1}z^*_{by}, \sigma_b^{-1}g_b(y+1;z,\theta), c_1^+, c_2^+\right).\end{aligned}$$

2. $g_b(y;z,\theta) \leq g_b(y+1;z,\theta) \leq z^*_{by}$.

$$\begin{aligned}\wp_{by}(z,\zeta) &= \sigma_b^{-1}\int_{g_b(y;z,\theta)}^{g_b(y+1;z,\theta)} \Phi\left(\frac{m^-_{by}(z,v,\theta)-\rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv \\ &= \tilde{\Delta}\left(\sigma_b^{-1}g_b(y;z,\theta), \sigma_b^{-1}g_b(y+1;z,\theta), c_1^-, c_2^-\right).\end{aligned}$$

3. $z^*_{by} \leq g_b(y;z,\theta) \leq g_b(y+1;z,\theta)$.

$$\begin{aligned}\wp_{by}(z,\zeta) &= \sigma_b^{-1}\int_{g_b(y;z,\theta)}^{g_b(y+1;z,\theta)} \Phi\left(\frac{m^+_{by}(z,v,\theta)-\rho\frac{\sigma_d}{\sigma_b}v}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{v}{\sigma_b}\right)dv \\ &= \tilde{\Delta}\left(\sigma_b^{-1}g_b(y;z,\theta), \sigma_b^{-1}g_b(y+1;z,\theta), c_1^+, c_2^+\right).\end{aligned}$$

Using indicators for whether $z^*_{by} < g_b(y+1;z,\theta)$ and $z^*_{by} > g_b(y;z,\theta)$ to cover each of these cases gives

$$\wp_{by}(z,\zeta) = \begin{pmatrix} 1\left[z^*_{by} < g_b(y+1;z,\theta)\right]\tilde{\Delta}\left(\sigma_b^{-1}\max\left\{z^*_{by}, g_b(y;z,\theta)\right\}, \sigma_b^{-1}g_b(y+1;z,\theta), c_1^+, c_2^+\right) \\ +1\left[z^*_{by} > g_b(y;z,\theta)\right]\tilde{\Delta}\left(\sigma_b^{-1}g_b(y;z,\theta), \sigma_b^{-1}\min\left\{z^*_{by}, g_b(y+1;z,\theta)\right\}, c_1^-, c_2^-\right) \end{pmatrix}. \tag{A.23}$$

This produces (5.3) by noting that the variables defined in (5.5) and (5.6) satisfy

$$\begin{aligned} m_1^+ &= \frac{c_1^+}{\sqrt{\left(c_2^+\right)^2+1}}, \qquad m_2^+ = -\frac{c_2^+}{\sqrt{\left(c_2^+\right)^2+1}}, \\ m_1^- &= \frac{c_1^-}{\sqrt{\left(c_2^-\right)^2+1}}, \qquad m_2^- = -\frac{c_2^-}{\sqrt{\left(c_2^-\right)^2+1}}, \end{aligned}$$

from which it follows from (A.21) and (A.22) that $\wp_{by}(z,\zeta)$ in (A.23) is equal to (5.3) in the statement of the Proposition for each b and $y \in \{1,2\}$. ■

B Data

In the application in Section 6 we used data on purchases of women's razor blades for the years 2004-2005 in Great Britain. The razor blade market is divided into three different sectors: cartridges bought with a razor (referred to as "system razors"), cartridges bought alone (referred to as "system blades"), and disposable razors. The original data consists of 7234 observations in which the main shopper was a female. Table 7 shows the observed market share of the three sectors in 2004-2005.

Sector	Total
System razors	16.80%
System blades	33.67%
Disposable razors	49.53%

Table 7: Market shares of the sectors System razors, System blades and Disposable razors in 2004-2005

For the application we concentrate on the market for system blades, where consumers buy a set of cartridges to use with a handle they already own from Gillette or Wilkinson Sword. We define the outside option as the purchase of a disposable razor. On average in our sample, using equation (6.1) for the per-cartridge unit price of system blades, a double blade cartridge costs £0.79 and a triple blade cartridge costs £1.45. Table 8 provides the average unit price per cartridge for each (b, y) combination of cartridges in our sample as estimated using equation (6.1).

As discussed in Section 2, Restriction A4 requires the explanatory variables and the unobservables to be stochastically independent. In our application prices are included as exogenous variable, so the independence assumption requires price to be independent of unobservable heterogeneity. Given the relatively small cost of razor blades as indicated in Table 8 it seems reasonable to assume that this cost makes up only a small fraction of total expenditure, such that consumers are unlikely to choose where to shop for their groceries and personal care items on the basis of razor blade prices. Thus we think the assumption of price exogeneity is reasonable in this context.

Trading Company	Blade Type	
	Double Blade	Triple Blade
Gillette	£0.75	£1.32
	(0.0786)	(0.1028)
Wilkinson Sword	£0.85	£1.86
	(0.0421)	(0.2398)

Table 8: Average unit prices of system blade cartridges in 2004-2005. Standard deviations in parentheses.

The women's market for reusable razors for the years 2004-2005 is dominated by two firms, Gillette and Wilkinson Sword, each offering razors and cartridges with two or three blades.[21] Gillette's twin-blade reusable razor, Sensor for Women, was launched in 1992 and the first three-blade reusable razor, Venus (Original) Razor, was introduced in 2001. Wilkinson Sword introduced the double-edged system razor, Lady Protector, in 1994, and the triple blade reusable razor for women, Intuition, in 2003.[22]

We use households in which the main shopper of the household is a female between 18-50 years old who is active in the labor force. This includes women who work full time, work part time, are unemployed or not working, or in full time education.[23] In the analysis we also

[21]Non-Gillette or non-Wilkinson Sword cartridges for reusable razors, for example stores' own-label system blades, were dropped from the sample as they accounted for only a small percentage of the market share for system blades in our data.

[22]Datta (2019) and Women's Razors, Shavers & Shaving Products UK | Wilkinson Sword (source: www.wilkinsonsword.com/en-gb/womens/). In 1992 Schick-Wilkinson Sword was formed and the Schick name was use in North America and elsewhere, while the Wilkinson Sword name was used in Europe (History of Wilkinson Sword Ltd. – FundingUniverse, source: www.fundinguniverse.com/company-histories/wilkinson-sword-ltd-history/.)

[23]Retired individuals were excluded from the sample. In our analysis "Employed" corresponds to working some hours while "Unemployed" corresponds to either unemployed/not working or full time education.

include the marital status of the main shopper,[24] and a variable indicating whether there is more than one female in the household. The sample used in our analysis consists of 4842 observations. Table 9 gives summary statistics of the main shopper characteristics.

Employment status		Marital status		No of females	
Works more than 30 hours	39.20%	Married	61.83%	One female	43.04%
Works 8-29 hours	29.18%	Single	31.50%	More than one	56.96%
Works less than 8 hours	2.38%	Divorced/Widowed/Separated	6.67%		
Unemployed/not working	28.13%				
Full time education	1.12%				

Table 9: Main shopper characteristics for 2004-2005.

For each observation in the sample we observe whether they purchased cartridges for reusable razors or disposable razors and the type of blade they bought, as well as the total amount they spent, the pack size of the product they bought, the month they made the purchase, and the store in which the purchase was made. As shown in Table 10, cartridges of system blades were observed in pack sizes of 3-8 cartridges, with double blade cartridges only offered in a pack size of 5. In the calculation of the average price in equation (6.1) and of the counterfactual prices in equation (6.2) the pack sizes were redefined as small (S) if they contained 3 or 4 cartridges/disposable razors, medium (M) if they contained 5 or 6 cartridges/disposable razors, and large (L) if they contained 8 or more cartridges/disposable razors, as shown in Table 11.

Pack size	System Blades				Disposables
	Double Blade	Triple Blade	Gillette	Wilkinson Sword	
3		22.00%		59.92%	0.10%
4		72.80%	72.65%		23.74%
5	100.00%		24.24%	34.40%	43.67%
6		3.05%	0.97%	5.67%	3.62%
8		2.15%	2.15%		13.21%
more than 8					15.65%

Table 10: Pack sizes observed.

The counterfactual prices in equations (6.2) and (6.3) were calculated by conditioning and not conditioning on the pack size of the product purchased, respectively. As is evident from Table 10 not all blade types and not all brands are observed being offered in all pack sizes. Table 13 gives the estimates of regressions (6.2) and (6.3). For the calculation of the

[24]The married indicator is zero for those who indicated they were single, divorced, widowed, or separated.

Pack size	System Blades	Total
S	69.23%	42.34%
M	29.19%	39.92%
L	1.57%	17.74%

Table 11: Pack size grouping.

average price in equation (6.1) and of the counterfactual prices in equations (6.2) and (6.3), the stores were grouped according to Table 12.

Company group	Percent of observed purchases
Asda	22.24%
Boots	8.57%
Co-op	0.50%
Kwik Save	0.56%
Morrisons	7.95%
Safeway	1.47%
Sainsbury's	8.01%
Savacentre	0.64%
Somerfield	0.91%
Superdrug	4.03%
Tesco	25.01%
Waitrose	0.33%
Wilkinsons	13.78%
Default[25]	3.61%
All other	2.40%

Table 12: Company groups of stores observed.

C Alternative Specific Conditional Logit Model

In this Appendix the results in Section 6.3 are compared to the standard alternative specific conditional logit (asclogit) model. Following McFadden (1974) choice probabilities of the

[25]Before grouping the stores in the 15 categories in Table 12, the different store names were identified using the shop identifier and/or the shop code available from the purchase data. The "Default" category in the original store data contains different shop codes. One of these shop codes also corresponded to the shop code of some "Tesco" stores in our data. These observations were re-classified from the "Default" to the "Tesco" category. All the other shop codes in the "Default" category remained in the category.

	Specification (6.2)	Specification (6.3)
Wilkinson Sword	0.4048	0.3961
	(0.0080)	(0.0081)
Triple blade	0.5043	0.6948
	(0.0240)	(0.0082)
Medium pack size	-0.1988	
	(0.0236)	
Large pack size	-0.0205	
	(0.0277)	
Constant	0.7385	0.5456
	(0.0429)	(0.0370)

Notes: Monthly and store dummies are suppressed. Standard errors in parentheses.

Table 13: OLS estimates of (6.2) and (6.3) conditional on system blade purchases.

asclogit model are specified as:

$$\wp_{by}(z;\zeta) = \frac{\exp(p_{by}\gamma_1 + 1[y=2]\gamma_2 + 1[b=2]\gamma_3 + x\beta_{by})}{\sum_{(b',y')\in\mathcal{M}_{BY}} \exp(p_{b'y'}\gamma_1 + 1[y'=2]\gamma_2 + 1[b'=2]\gamma_3 + x\beta_{b'y'})},$$

where x denotes the same individual characteristics as in our model and p_{by} is the price for each of the four system blades alternatives each individual would have faced at the time of purchase. The specification includes both brand and high-quality (triple blade) product-specific dummy variables $1(b = 2)$ and $1(y = 2)$, respectively.[26] The mean value of the utility of the outside good, disposable razors, is imposed to be zero by setting the value of each of the alternative specific covariates, i.e. the price and the quality and brand dummies, for option $(0, 0)$ to zero. The estimated coefficients using only specification (6.2) are given in Table 14. These results were then used to calculate the predicted own and cross price elasticities, which following e.g. Cameron and Trivedi (2009) take the form:

$$\eta_{ibykl} = \begin{cases} \gamma_1[1-\wp_{by}(z_i;\beta,\gamma)]p_{iby} & \text{if } (b,y)=(k,l), \\ -\gamma_1\wp_{kl}(z_i;\beta,\gamma)p_{ikl} & \text{if } (b,y)\neq(k,l). \end{cases}$$

Table 15 gives the estimates of the average and the 0.2, 0.5 and 0.8 quantiles of household own- and cross-price elasticities. For example, the mean own price elasticity of a Gillette double blade cartridge, $(1, 1)$, is -0.8871, while the mean cross-price elasticity of the proba-

[26]Constant terms are excluded from the alternative-specific utilities as these would not be identified with the inclusion of the brand and quality dummy variables.

	Specification (6.2)				
	γ	β_{11}	β_{12}	β_{21}	β_{22}
Price	-1.2550				
	(0.1250)				
Triple Blade	0.5209				
	(0.1582)				
Wilkinson Sword	-1.3582				
	(0.1651)				
Age31-40		-0.2304	-0.1440	-0.3202	0.2973
		(0.1172)	(0.0854)	(0.1831)	(0.1443)
Age41-50		-0.6836	-0.4300	-0.1935	0.1617
		(0.1567)	(0.1014)	(0.2013)	(0.1640)
Married		-0.1051	0.1680	0.1166	0.2867
		(0.1066)	(0.0769)	(0.1589)	(0.1259)
Employed		-0.3358	0.3436	0.1396	0.8254
		(0.1034)	(0.0812)	(0.1558)	(0.1416)
Females		-0.7164	-0.0819	0.2128	-0.2498
		(0.1077)	(0.0754)	(0.1572)	(0.1199)

Notes: Standard errors in parentheses.

Table 14: Alternative specific conditional logit regression estimates.

bility of choosing any other alternative with respect to the price of the Gillette double blade is 0.0874.

D Monte Carlo Experiments

In this Appendix we report the results of Monte Carlo experiments to compare the finite sample performance of Wald and profile QLR confidence intervals for inference with our model.[27] For these experiments we generated data from the partially ordered probit model, with the number of parameters matching those employed in the subsequent application. There were five individual-specific dummy variables with corresponding coefficients $\beta_{b1}, ..., \beta_{b5}$ for each $b = 1, 2$. Each product offering had a price P_{by} generated differently in each of the three data generation processes (DGPs) – referred to as DGP1, DGP2, and DGP3 – as described below. The linear-in-price specification described by (4.6) was used.

[27]Code for the Monte Carlo experiments is available at `https://sites.google.com/site/amr331/home/por-code`.

	Own Price Elasticities			Cross Price Elasticities		
	Specification (6.2)					
$(\mathbf{1},\mathbf{1})$						
mean		-0.8871			0.0874	
quantiles	-1.0054	-0.9074	-0.7554	0.0514	0.0797	0.1207
$(\mathbf{1},\mathbf{2})$						
Mean		-1.2300			0.3583	
Median	-1.3569	-1.2362	-1.0941	0.3131	0.3563	0.4050
$(\mathbf{2},\mathbf{1})$						
mean		-1.4073			0.0562	
quantiles	-1.5371	-1.4531	-1.2659	0.0489	0.0546	0.0685
$(\mathbf{2},\mathbf{2})$						
mean		-1.9675			0.1478	
quantiles	-2.1028	-1.9759	-1.8036	0.0940	0.1550	0.1952

Table 15: Alternative specific conditional logit mean and 0.2, 0.5, and 0.8 quantile estimates of household elasticities.

To simulate data population parameter values were set as follows.

$$\begin{aligned} &\gamma_1 = 1, \quad \gamma_2 = 0.8, \quad \delta_1 = -1.5, \quad \delta_2 = -1.2, \quad \rho = 0.5, \quad \sigma = 1, \\ &\beta_1 = (1.3, 0.3, -0.1, -0.3, 0.7)', \quad \beta_2 = (1.0, 0.3, -0.1, -0.3, 0.7)'. \end{aligned} \tag{D.1}$$

In our application the first two components of X, X_1 and X_2, are dummy variables indicating whether age of a female shopper is from 31-40, or 41-50, with 18-30 denoting the base category. These variables were drawn such that $\Pr[X_1 = 1] = 0.426$, $\Pr[X_2 = 1] = 0.234$, and $X_1 + X_2 \leq 1$. The remaining components of X are dummy variables for marriage, employment, and a variable "more_females" indicating the presence of more than one female in the household. These were generated from the Bernoulli distribution with parameters 0.4, 0.85, and 0.554, respectively. The age variable from which X_1 and X_2 were generated and the remaining dummy variables were drawn independently of one another.

Prices $(P_{11}, P_{12}, P_{21}, P_{22})$ were generated independently of X, as follows. First, for each DGP and for each observation a vector ε was drawn from the bivariate normal distribution with each component having mean zero and variance one, with correlation 0.25. In DGP1 prices P_{11} and P_{21} were generated independently, and uniformly on the intervals $[1, 4]$ and $[1.35, 2.15]$, respectively. Prices P_{12} and P_{22} were then set to $P_{12} = P_{11} + \varepsilon_1$ and $P_{22} = P_{21} + \varepsilon_2$. In this DGP, prices P_{12} and P_{22} thus both have positive density on all of $\mathbb{R}$ conditional on all other variables. This implies that there is positive probability that the price of the higher quality product for either brand b undercuts the price of the lower quality product, i.e. $P_{b2} < P_{b1}$, as could happen under a promotion for the higher quality product. In such cases the conditional probability of choosing the lower quality product for the brand will be zero. Moreover, the large support for both P_{12} and P_{22} imply that this happens with positive probability for both brands, in which case the choice problem reduces to a simple multinomial choice setting between each brand's higher quality product and the outside option. Thus, the large support of these variables, artificial though it may be, demonstrates a setting in which point identification may be achieved. This is borne out in the Monte Carlo simulations below.

In practice prices will not have support on the entire real line, and neither DGP2 nor DGP3 have this feature. In DGP2, P_{11} and P_{21} were generated independently from the uniform distribution on $[1, 2]$ and $[1.35, 2.15]$, respectively, and each P_{b2} was set to $P_{b1} + \max\{1, \min\{|\varepsilon_b|, 2\}\}$. Thus the higher quality product for each brand always has a higher price than the lower quality product of that brand. Moreover, all prices have continuous, but bounded support. In DGP3, P_{11} and P_{21} were generated the same way, but the term added on to P_{b1} to determine P_{b2} was instead rounded to the nearest integer (which was either one or two) before adding. In this design prices again have bounded continuous support, but for each b the conditional support of P_{b2} given P_{b1} is discrete.

With variables X and prices $P = (P_{11}, P_{12}, P_{21}, P_{22})$ generated as described above, and unobservables $V = (V_1, V_2)$ drawn from the bivariate normal distribution with parameters ρ and σ, data (b_i, y_i, x_i, p_i) were generated with each (b_i, y_i) solving the individual choice problem with the corresponding (x_i, p_i, v_i) and utility parameters as in (D.1). The expression (5.3) obtained for choice probabilities in Proposition 2 was used in the log-likelihood function based on n observations in each experiment, with $n \in \{200, 500, 1000, 2000\}$. In preliminary investigations, choice probabilities computed using (5.3) conditional on several values of observable variables were compared to those obtained using the integral formula (5.1) and those obtained by simulation, and these were all found to be in close agreement up to negligible computation difference.

In order to compare the empirical coverage frequencies of the Wald and QLR confidence intervals in each repetition of our Monte Carlo simulations, we carried out the following steps. First, the R package Ghalanos and Stefan (2015) was used to minimize $-L_n(\zeta)$ with respect to the full parameter vector ζ, producing an optimizing vector $\hat{\zeta}_{ML}$ and an optimal value L_n^*. To ensure accuracy 500 randomly generated starting values were employed using the function gosolnp.[28] The optimization routine also returned a numerical approximation to the Hessian at the optimal value, and this was used to compute asymptotic variance estimators $\widehat{var}(\hat{\zeta}_k)$ for each component of the maximum likelihood estimator $\hat{\zeta}_{ML}$. There is no guarantee that ζ is point identified. If it is point identified confidence intervals for each parameter component based on the usual asymptotic normal approximation should be expected to perform well, but if it is not point identified the classical theory will be invalid. Nonetheless QLR confidence intervals can sometimes remain valid under partial identification as shown by Chen, Christensen, and Tamer (2018).[29]

In Monte Carlo experiments where the true population parameter is known, the same routine was also used to compute the maximum likelihood estimator taking the values of ρ and σ fixed at their population values. In our application ρ and σ are not known, so this approach is infeasible. However, with these parameters known, the rest of the parameters are point identified under mild conditions on the variation in observable payoff shifters. Thus, these confidence intervals should be expected to perform well, and in our Monte Carlo experiments this was indeed the case. In our reported results, we refer to confidence intervals for each ζ_k based on the maximum likelihood estimator $\hat{\zeta}_k$ plus or minus $1.96 \times \sqrt{\widehat{var}(\hat{\zeta}_k)}$ as Wald confidence intervals, considering both cases where maximum likelihood was carried out with ρ and σ fixed at their population values, as well as with ρ and σ as additional parameters to estimate. Only the latter approach is feasible if the population values of ρ and σ are unknown. In order to compute Monte Carlo coverage frequencies of $\mathcal{C}^{LR}_{\alpha,k}$ for ζ_k we additionally computed $Q_{k,n}(\zeta_k)$ as defined in (6.5) at the population value of ζ_k and checked whether $Q_{k,n}(\mu) \leq \chi^2_{1,\alpha}$ in each simulation.[30]

The empirical coverage frequency of the three different procedures for DGPs 1-3 out of 1000 Monte Carlo repetitions for each sample size are reported in Tables 16, 17, and 18. The target coverage level in each case was 0.95. All procedures performed reasonably well, although coverage probabilities for the distributional parameters are substantially below the

[28]In Monte Carlo simulations the population parameter value was also used as an additional starting value. The number of randomly generated starting values was chosen based on experimentation; increasing it further was not found to be beneficial.

[29]All reported coverage frequencies are for individual parameters ζ_k. Chen, Christensen, and Tamer (2018) provide sufficient conditions for the QLR confidence interval to achieve asymptotic coverage for the identified set for ζ_k, which also guarantee at least as high a level of asymptotic coverage for ζ_k.

[30]Our implementation employed the solnp function from the package Ghalanos and Stefan (2015) to compute $\sup_{\zeta \in \Upsilon : \zeta_k = \mu} L_n(\zeta)$. In the constrained optimizations conducted with solnp, the population value of ζ_{-k} was used as a starting value to speed up computations. In terms of coverage frequency for ζ_k, this was found to produce the same results in a subset of the Monte Carlo iterations attempted when as many of 500 random starting values were used.

<u>**DGP 1**</u>	**Realized Coverage Probability,** $n = 200, 500, 1000, 2000$											
Parameter	**Wald (ρ, σ known)**				**Wald (ρ, σ unknown)**				**QLR**			
γ_1	.935	.938	.954	.950	.952	.951	.950	.944	.929	.941	.947	.935
γ_2	.944	.954	.950	.961	.934	.956	.951	.937	.938	.949	.947	.934
δ_1	.927	.952	.951	.958	.937	.960	.963	.954	.943	.955	.953	.950
δ_2	.940	.949	.954	.957	.907	.945	.947	.939	.932	.948	.941	.937
β_{11}	.957	.952	.951	.949	.951	.956	.946	.950	.942	.948	.939	.942
β_{12}	.947	.961	.944	.949	.950	.963	.955	.945	.933	.954	.949	.933
β_{13}	.959	.953	.951	.949	.961	.961	.953	.950	.947	.956	.944	.943
β_{14}	.954	.948	.959	.952	.962	.950	.960	.953	.942	.941	.959	.946
β_{15}	.954	.956	.957	.952	.949	.968	.961	.956	.945	.962	.956	.948
β_{21}	.960	.963	.958	.957	.922	.943	.938	.942	.938	.946	.936	.941
β_{22}	.942	.956	.947	.953	.938	.961	.961	.952	.931	.950	.947	.941
β_{23}	.953	.949	.961	.959	.969	.955	.964	.963	.951	.946	.951	.951
β_{24}	.964	.946	.941	.962	.930	.940	.926	.951	.945	.937	.931	.942
β_{25}	.957	.957	.956	.961	.924	.952	.957	.941	.955	.955	.953	.938
ρ	–				.899	.928	.934	.944	.933	.940	.940	.941
σ	–				.908	.941	.956	.927	.942	.947	.957	.923

Table 16: Monte Carlo coverage frequencies out of 1000 simulations for sample sizes $n = 200, 500, 1000, 2000$ for DGP1, as described in the text.

<u>**DGP 2**</u>	**Realized Coverage Probability,** $n = 200, 500, 1000, 2000$											
Parameter	**Wald** (ρ, σ **known**)				**Wald** (ρ, σ **unknown**)				**QLR**			
γ_1	.956	.956	.957	.944	.936	.961	.958	.942	.936	.951	.955	.939
γ_2	.955	.947	.954	.948	.836	.905	.926	.943	.899	.935	.942	.948
δ_1	.962	.960	.955	.957	.923	.957	.958	.963	.939	.963	.956	.959
δ_2	.961	.945	.951	.950	.812	.897	.923	.940	.892	.921	.941	.942
β_{11}	.942	.962	.953	.955	.928	.947	.942	.952	.939	.958	.946	.959
β_{12}	.942	.950	.953	.944	.951	.953	.951	.942	.943	.953	.953	.942
β_{13}	.953	.950	.951	.956	.951	.952	.952	.950	.948	.949	.949	.949
β_{14}	.949	.950	.946	.963	.961	.956	.948	.957	.949	.955	.947	.956
β_{15}	.940	.958	.938	.957	.927	.938	.934	.949	.934	.952	.938	.948
β_{21}	.942	.956	.958	.941	.852	.905	.926	.934	.912	.939	.945	.937
β_{22}	.948	.957	.947	.941	.900	.937	.940	.946	.927	.947	.945	.943
β_{23}	.931	.960	.947	.935	.949	.966	.959	.952	.927	.954	.951	.943
β_{24}	.946	.939	.947	.962	.894	.933	.932	.962	.925	.947	.940	.953
β_{25}	.940	.948	.941	.941	.866	.915	.932	.938	.925	.933	.942	.936
ρ	–				.710	.840	.885	.904	.899	.929	.938	.942
σ	–				.843	.910	.928	.955	.896	.929	.940	.950

Table 17: Monte Carlo coverage frequencies out of 1000 simulations for sample sizes $n = 200, 500, 1000, 2000$ for DGP2, as described in the text.

DGP 3	Realized Coverage Probability, $n = 200, 500, 1000, 2000$											
Parameter	**Wald (ρ, σ known)**				**Wald (ρ, σ unknown)**				**QLR**			
γ_1	.948	.958	.939	.952	.954	.964	.948	.950	.947	.958	.947	.947
γ_2	.949	.955	.945	.943	.911	.930	.951	.950	.929	.946	.944	.952
δ_1	.954	.950	.950	.956	.955	.959	.956	.958	.951	.963	.956	.960
δ_2	.960	.944	.948	.948	.907	.932	.953	.953	.923	.931	.957	.941
β_{11}	.936	.945	.948	.952	.937	.953	.958	.954	.938	.954	.958	.951
β_{12}	.949	.950	.956	.948	.956	.954	.959	.940	.952	.948	.953	.937
β_{13}	.948	.936	.954	.951	.955	.944	.950	.950	.951	.941	.947	.947
β_{14}	.939	.953	.954	.956	.940	.950	.954	.960	.935	.950	.951	.955
β_{15}	.941	.952	.943	.946	.941	.957	.954	.959	.944	.953	.952	.956
β_{21}	.938	.945	.953	.954	.917	.948	.953	.943	.934	.953	.950	.940
β_{22}	.942	.956	.943	.948	.947	.961	.952	.947	.946	.956	.950	.951
β_{23}	.945	.959	.942	.944	.968	.965	.959	.946	.931	.953	.942	.944
β_{24}	.948	.945	.950	.963	.923	.950	.943	.960	.936	.948	.944	.961
β_{25}	.938	.945	.946	.948	.931	.933	.949	.947	.950	.941	.954	.946
ρ	–				.795	.897	.925	.923	.916	.943	.947	.937
σ	–				.918	.939	.957	.949	.927	.932	.950	.943

Table 18: Monte Carlo coverage frequencies out of 1000 simulations for sample sizes $n = 200, 500, 1000, 2000$ for DGP3, as described in the text.

nominal level at smaller sample sizes. This is particularly so for the Wald procedure in which ρ and σ are treated as unknown parameters, while under-coverage from the QLR procedure is less severe at smaller sample sizes.

References

Andrews, D. W. K. (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67(6), 1341–1383.

Aradillas-Lopez, A., and A. M. Rosen (2022): "Inference in Ordered Response Games with Complete Information," *Journal of Econometrics*, 226(2), 451–476.

Azzalini, A., and A. Genz (2016): *The R package* `mnormt`*: The multivariate normal and t distributions (version 1.5-4).*

Barseghyan, L., M. Coughlin, F. Molinari, and J. Teitelbaum (2021): "Heterogeneous Choice Sets and Preferences," *Econometrica*, 89(5), 2015–2048.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Berry, S., and A. Pakes (2007): "The Pure Characteristics Demand Model," *International Economic Review*, 48, 1193–1225.

Bowden, R. (1973): "Theory of Parametric Idenification," *Econometrica*, 41(6), 1069–1074.

Bresnahan, T. F. (1987): "Competition and Collusion in the American Automobile Industry: the 1955 Price War," *Journal of Industrial Economics*, 35(4), 457–482.

Cameron, A. C., and P. K. Trivedi (2009): *Microeconometrics using stata.* College Station (TX), Stata Press.

Chen, X., T. M. Christensen, and E. Tamer (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86(6), 1965–2018.

Chen, X., E. Tamer, and A. Torgovitsky (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," Cowles Foundation working paper 1836.

Chesher, A., and A. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85(3), 959–989.

Chesher, A., and A. M. Rosen (2020): "Econometric Modeling of Interdependent Discrete Choice with Applications to Market Structure," CeMMAP working paper CWP25/20.

Chesher, A., A. M. Rosen, and Z. Siddique (2019): "Estimating Endogenous Effects on Ordinal Outcomes," CeMMAP working paper CWP66/19.

Chesher, A., A. M. Rosen, and K. Smolinski (2013): "An Instrumental Variable Model of Multiple Discrete Choice," *Quantitative Economics*, 4(2), 157–196.

Datta, Y. (2019): "The U.S. Women's Razor-Blade Market: A Competitive Profile," *Journal of Economics and Public Finance*, 5, 491–508.

Davies, S., C. Waddams, and C. Wilson (2009): "How Far Does Economic Theory Explain Competitive Nonlinear Pricing in Practice?," Centre for Competition Policy Working Paper 09-7.

Dubin, J., and D. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345–362.

Ghalanos, A., and T. Stefan (2015): *Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method (version 1.16).*

Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63(4), 891–951.

Hausman, J., and D. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46(2), 403–426.

Kenkel, B. (2015): *The R package `pbivnorm`: Vectorized Bivariate Normal CDF (version 0.60).* Based on the `mnormt` package. Fortran code by Alan Genz.

Liu, X., and Y. Shao (2003): "Asymptotics for Likelihood Ratio Tests under Loss of Identifiability," *Annals of Statistics*, 31(3), 807–832.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press.

Milgrom, P., and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157–180.

Nevo, A. (2000): "A practitioner's guide to estimation of random-coefficients logit models of demand," *Journal of economics & management strategy*, 9(4), 513–548.

Owen, D. B. (1980): "A Table of Normal Integrals," *Communications in Statistics - Simulation and Computation*, 4(4), 389–419.

Prekopa, A. (1973): "On logarithmic concave measures and functions," *Acta Sci. Math.(Szeged)*, 34, 335–343.

Redner, R. (1981): "Note on the Consistency of the Maximum Likelihood Estimate for Nonidentifiable Distributions," *Annals of Statistics*, 9(1), 225–228.

Self, S. G., and K.-Y. Liang (1987): "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions," *Journal of the American Statistical Association*, 82(398), 605–610.

Song, M. (2015): "A Hybrid Discrete Choice Model of Differentiated Product Demand with an Application to Personal Computers," *International Economic Review*, 56(1), 265–301.