cemmap

centre for microdata methods and practice

A discrete choice model for partially ordered alternatives

Eleni Aristodemou
Adam M. Rosen

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP62/19

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

A Discrete Choice Model for Partially Ordered Alternatives.*

Eleni Aristodemou†
University of Amsterdam

Adam M. Rosen‡
Duke University, CeMMAP and IFS

November 29, 2019

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 a parametric generalization of the ordered probit model. Conditions for point identification are established when the distribution of unobservable heterogeneity is known, but remain elusive when the

*We thank Tim Christensen, Francesca Molinari, Lars Nesheim and several conference and seminar audiences for helpful discussions and comments. Khuong (Lucas) Do 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 from the European Research Council (ERC) under grants ERC-2009-StG-240910-ROMETA and ERC-2009-AdG, grant agreement number 249529. Data supplied by TNS UK Limited. The use of TNS UK Ltd. data in this work does not imply the endorsement of TNS UK Ltd. in relation to the interpretation or analysis of the data. All errors and omissions remain the responsibility of the authors.

†Address: Quantitative Economics Section, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1001 NJ, Amsterdam, Netherlands. Email: e.aristodemou@uva.nl.

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

distribution is instead restricted to the multivariate normal family with parameterized variance. Rather than invoke the restriction that the distribution is known, or simply assume that model parameters are point identified, we consider the use of inference methods that allow for the possibility of set identification, and which are therefore robust to the possible lack of point identification. A Monte Carlo analysis is provided in which inference is carried out using a method proposed by Chen, Christensen, and Tamer (2018), which is insensitive to the possible lack of point identification and is found to perform adequately. An empirical illustration is then conducted using consumer purchase data in the UK to study consumers' choice of razor blades 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 $q = 1, 2, ..., \overline{q}_b$. Individuals are assumed to have ordered preferences over the vertical quality dimension, *within* each horizontally differentiated category, but preferences *across* horizontal categories are unordered. Prominent examples of product offerings in which different firms compete to sell vertically differentiated products to consumers include airline tickets for a given city pair where each airline offers vertically differentiated travel classes, and vertically-differentiated cable television packages offered by multiple providers.

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. Yet 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).

We combine features of models for ordered and unordered choice in order to incorporate both aspects of differentiation. Relative to existing methods, this approach allows the model to respect 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.

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, potentially, multiple equilibria, as encountered for instance in simultaneous equations model for ordered actions considered by Aradillas-Lopez and Rosen (2014).

Nonetheless, despite the lack of strategic interactions in the single-agent decision problem studied here, the discrete nature of both dimensions of the decision problem and the unordered-ordered nature of the brand-quality decision gives rise to identification challenges. This is true even when we impose a linear index structure to model within-brand utility. We first show that under a rank condition there is point identification of model parameters if the distribution of unobserved heterogeneity is known and log-concave, for example multivariate normal. If instead unobserved heterogeneity is assumed to belong to a family of distributions with parameters indexed by $\Sigma \in \mathbf{\Lambda}$, this identification result no longer holds. We show however that under some mild conditions the identified set for utility parameters θ takes the form $\Theta^* = \{\theta(\Sigma) : \Sigma \in \mathbf{\Lambda}\}$ where $\theta(\Sigma)$ maximizes the expected log-likelihood for the heterogeneity distribution with parameters Σ.

More generally, since the model pertains to a single agent decision problem it is complete, and a log-likelihood can be constructed. The identified set for the combined vector of all model parameters (i.e. both payoff and distributional parameters (θ, Σ)), though possibly not a singleton, can be characterized as the set of maximizers of the expected log-likelihood. This characterization in turn permits application of results on the distribution of likelihood ratio statistics when point identification need not hold, such as those of Liu and Shao (2003) for parametric likelihood models and Chen, Tamer, and Torgovitsky (2011) for semi-parametric likelihood models. In Monte Carlo experiments we investigate the use of an inference approach developed by Chen, Christensen, and Tamer (2018) – henceforth CCT – that allows construction of confidence intervals for individual parameters, and which is easy to implement. These confidence intervals are valid and slightly conservative when there is partial identification, but have the desirable feature that they are asymptotically exact if the parameter of interest is point identified. This inference method is found to perform well in our Monte Carlo experiments, and is subsequently used in an empirical application to consumer choice of vertically differentiated razors from each of two different brands using data on consumers in the United Kingdom.

The paper proceeds as follows. In Section 2 we provide our econometric model for partially ordered response in its most general form. In Section 3 we provide identification analysis. In Section 4 we provide a parametric probit-type model with two brands each with two quality levels, and illustrate how characterization of the identified set simplifies in this context. Section 5 provides details for computation of the log-likelihoods, implementation of results from CCT for construction of confidence intervals, and Monte Carlo analysis illustrating performance of the inference approach. Section 6 presents an application to the market for women's razor blades using consumer data from the United Kingdom in the early 2000s. Section 7 concludes and discusses directions of continuing research. Proofs of propositions and theorems are provided in the Appendix.

2 The Model

Each individual in the population is characterized by observables (Y, B, X) 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 $B \in \mathcal{B} \equiv \left\{1, ..., \bar{b}\right\}$, or an outside alternative denoted by $Y = 0$. The set $\overline{\mathcal{M}}_{BY} \equiv \{(b, y) : B \in \mathcal{B},\ Y \in \mathcal{Y}_b\}$ denotes the set of possible (B, Y) alternatives, and $\mathcal{M}_{BY} \equiv \overline{\mathcal{M}}_{BY} \cup \{(b, 0) : B \in \mathcal{B}\}$ denotes the joint support of (B, Y). When $Y = 0$ brand choice is undefined, and any (b, y) pair with $y = 0$ denotes the outside alternative. The set $\mathcal{X}$ denotes the support of observable covariates X, such as individual characteristics. The vector $V \in \mathbb{R}^{\bar{b}}$ represents unobserved heterogeneity that affects individuals' preferences both within and across types through the utility specification now described. The probability measure of V is denoted $G(\cdot)$ so that for any set $\mathcal{S} \subseteq \mathbb{R}^{\bar{b}}$, $G(\mathcal{S}) \equiv \Pr[V \in \mathcal{S}]$.

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

$$U_{by} \equiv u(b, y, x, v_b), \tag{2.1}$$

where each function $u(b, y, x, v_b)$ is strictly increasing in v_b for each $(b, y, x) \in \overline{\mathcal{M}}_{BY} \times \mathcal{X}$. Each individual chooses precisely one (B, Y) pair. We normalize the utility from the outside alternative $(Y = 0)$ to zero and define

$$U_{b0} \equiv 0,$$

for each $b \in \mathcal{B}$. Because the choice $Y = 0$ is intended as an outside alternative, this notation will prove convenient when comparing utilities of the general form U_{by}, but the value of b when $y = 0$ carries no meaning.

We assume that each individual chooses the alternative that maximizes her utility.[1] For any $b \in \mathcal{B}$ let

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

denote the indirect utility and optimal choice of Y, respectively, if the individual's alternatives were limited to only those of type b. 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 X, 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

$$B = \operatorname*{argmax}_{b \in \mathcal{B}} U_b^*, \qquad Y = Y_B^*. \tag{2.2}$$

Note in the case where none of the products deliver positive utility, $Y = 0$, so that $U_b^* = Y_b^* = 0$ for all $b \in \mathcal{B}$, and a purchase is not made from any brand.

Restriction A1: (Probability space) (B, Y, X, V) are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ contains the Borel sets. The support of B is $\mathcal{B} \equiv \{1, ..., \bar{b}\}$, the support of Y is $\mathcal{Y}_b \equiv \left\{0, 1, ..., \max_{b \in \mathcal{B}} \bar{y}_b\right\}$, and their joint support is denoted $\mathcal{M}_{BY}$. The support of (X, V) is $\mathcal{X} \times \mathcal{V}$ where $\mathcal{V} \subseteq \mathbb{R}^{|\mathcal{B}|}$.

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

Restriction A3: (Distribution of unobserved heterogeneity) The conditional distribution of V given $X = x$ is absolutely continuous with respect to Lebesgue measure with everywhere positive density on $\mathbb{R}^{|\mathcal{B}|}$.

Restriction A4: (Independence) X 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).

[1] Under Restriction A3 below ties in the utility obtained from different alternatives occur with zero probability conditional on any realization of x. 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 A6: (Utility maximization) Given (X, V), (B, Y) are chosen to maximize $u(B, Y, X, V_b)$, where u belongs to a known class of functions $\mathcal{U}$ satisfying (i) $u(b, 0, x, v_b) = 0$ for all (b, x, v_b), (ii) $u(b, y, x, v_b)$ is strictly increasing and continuous in v_b for all (b, y, x), and (iii) for each $(b, x) \in \mathcal{B} \times \mathcal{X}$, $\{u(b, y, x, v_b) : v_b \in \mathbb{R}\}$ satisfies the single-crossing property in (y, v_b), namely that if $v_b' > v_b$ and $y' > y$, then

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

Restriction A1 defines the underlying probability space and notation for the support of random variables (B, Y, X, V). Restriction A2 stipulates that the conditional distribution of (B, Y) given covariates x is point identified for almost every $x \in \mathcal{X}$, as would be 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 X and V. This is an important restriction. If X 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 a only small fraction of expenditure, such that individuals do not choose where to shop based on the price offered. It also holds if all individuals make their purchase decisions in a single market, where they face identical prices. 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 $f_x^0(b, y)$ for almost every $x \in \mathcal{X}$. The identification problem is to determine the set of structures that can generate the observed distributions $f_x^0(b, y)$. The set of structures $\mathcal{S}$ admitted by the model can be restricted to a parametric, semiparametric, or nonparametric class.

In particular, the underlying structure S maps to conditional distributions $f_x^0(b, y)$ through the specification of the individual choice problem. Restriction A6 specifies that individuals choose (B, Y) to maximize utility $u(B, Y, X, V_b)$, 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 by 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. This combined with the within brand monotonicity given by Restriction A6(ii) allows for characterizations of regions of unobservables that give rise to conditional choice probabilities for each brand-quality combination. This plays a key role in the identification of underlying structure S, as we show in the next Section.

3 Identification

We begin this section with a general characterization of the identified set of structures compatible with Restrictions A1-A6. We then show that if the model is correctly specified, the identified set can be written as the maximizers of the expected log-likelihood, and we derive the form of the multivariate integral delivering conditional choice probabilities as a function of the underlying structure S.

3.1 General Characterization of the Identified Set

Before adding further restrictions we first characterize the identified set of structures S under Restrictions A1-A6, denoted $\mathcal{S}_0(\mathcal{X})$. The notation expresses the dependence of the identified set on the support of the exogenous variables X. This set is by definition given by

$$\mathcal{S}_0(\mathcal{X}) \equiv \left\{(u, G) \in \mathcal{S} : \forall (b, y) \in \mathcal{M}_{BY},\, G(\mathcal{V}_{by}(x; u)) = f_x^0(b, y) \text{ a.e. } x \in \mathcal{X}\right\}, \tag{3.1}$$

where $\mathcal{V}_{by}(x; u)$ denotes that set of values for unobserved heterogeneity V on which (b, y) maximizes utility u:

$$\mathcal{V}_{by}(x; u) \equiv \left\{V \in \mathcal{V} : \forall \left(\tilde{b}, \tilde{y}\right) \neq (b, y), \quad u(b, y, x, v_b) \geq u\left(\tilde{b}, \tilde{y}, x, v_{\tilde{b}}\right)\right\}. \tag{3.2}$$

In words, $\mathcal{S}_0(\mathcal{X})$ is the set of admissible structures (u, G) that generate identified conditional choice probabilities $f_x^0(b, y)$ for all (b, y) and almost every $x \in \mathcal{X}$. Note that given the absolute continuity of the distribution of V and continuity of utility in unobserved heterogeneity, the sets $\mathcal{V}_{by}(x;u)$ and $\mathcal{V}_{\tilde{b}\tilde{y}}(x;u)$, $\left(\tilde{b},\tilde{y}\right) \neq (b,y)$, overlap at most on a set of Lebesgue measure zero, so that there is a unique utility maximizing pair (b,y) with probability one given any $x \in \mathcal{X}$. Hence $G(\mathcal{V}_{by}(x;u))$ is the conditional probability of observing (b,y) given $X = x$ when the utility function is u and $V \sim G$. Structures (u,G) that do not belong in the identified set $\mathcal{S}_0(\mathcal{X})$ in (3.1) are those such that the set

$$\mathcal{X}^*(u,G) \equiv \left\{x \in \mathcal{X} : \exists (b,y) \in \mathcal{M}_{BY} \text{ s.t. } G(\mathcal{V}_{by}(x;u)) \neq f_x^0(b,y)\right\}, \tag{3.3}$$

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

Given the representation of the identified set through the equalities $G(\mathcal{V}_{by}(x;u)) = f_x^0(b,y)$ 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 $\forall (b,y) \in \mathcal{M}_{BY}$ $G^*(\mathcal{V}_{by}(x;u^*)) = f_x^0(b,y)$ a.e. $x \in \mathcal{X}$.

This restriction requires that the distribution of (B,Y) conditional on X is obtained by at least one admissible structure $S^* \in \mathcal{S}$. This assumption is also imposed in Liu and Shao (2003) and Chen, Tamer, and Torgovitsky (2011), while Chen, Christensen, and Tamer (2018) note that their methods can be applied to perform inference on the identified set under misspecification in separable likelihood models. Nonetheless, interpretation of partially identifying models under misspecification is delicate, see Ponomareva and Tamer (2011), and this is not studied here.

Consider the expected log-likelihood function

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

where the expectation is taken with respect to population measure $\mathbb{P}$. It follows by arguments identical to those with singleton $\mathcal{S}_0(\mathcal{X})$ that $Q(u,G)$ attains its maximum at all $(u,G) \in \mathcal{S}_0(\mathcal{X})$, since by definition of $\mathcal{S}_0(\mathcal{X})$ these all produce the same probabilities $G(\mathcal{V}_{by}(x;u))$ for almost every x. 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*

$$\mathcal{S}_0(\mathcal{X}) \equiv \underset{(u,G)\in\mathcal{S}}{\text{argmax}}\ Q(u,G),$$

with $\mathcal{S}_0(\mathcal{X})$ *as defined in (3.1).*

In order to better understand the properties of the set $\mathcal{S}_0(\mathcal{X})$, we now investigate the form of the conditional choice probabilities $G(\mathcal{V}_{by}(x;u))$. Unless sufficiently strong parametric restrictions on $\mathcal{S}$ are imposed, $\mathcal{S}_0(\mathcal{X})$ may not be 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 $\mathcal{S}_0(\mathcal{X})$ 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 $G(\mathcal{V}_{by}(x;u))$ enables derivation of sufficient conditions for point identification, as well as computation of set estimates and inferential statistics when point identification fails.

3.2 Conditional Choice Probabilities

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

$$p_{by}(x;S) \equiv G(\mathcal{V}_{by}(x;u))$$

for brand-quality pair (b,y) given $X = x$, considered as a function of any structure $S = (u,G)$. Without parametric restrictions on u, the monotonicity and single-crossing conditions suffice to establish the representation of each choice probability $p_{by}(x;S)$ as a particular form of a $\bar{b}$-variate integral. Thus, given a specific (u,G), $p_{by}(x;S)$ can be computed by either numerical integration or simulation. The formal result follows.

Theorem 1 *Let Restriction A6 hold. Then for each* $(b,y,x) \in \mathcal{M}_{BY} \times \mathcal{X}$*, the region* $\mathcal{V}_{by}(x;u)$ *is a convex polytope in* $\mathbb{R}^{\bar{b}}$ *and the choice probability* $p_{by}(x;S)$ *takes the form*

$$p_{by}(x;S) = \int_{g_b(y)}^{g_b(y+1)} \int_{-\infty}^{h_{b,1}(y)} \cdots \int_{-\infty}^{h_{b,b-1}(y)} \int_{-\infty}^{h_{b,b+1}(y)} \cdots \int_{-\infty}^{h_{b,\bar{b}}(y)} dG(v), \qquad (3.4)$$

where $\{g_b(y) : y = (0, ..., \bar{y}_b + 1)\}$ *are within-brand threshold functions and* $\{h_{bk}(y) : k \neq b\}$ *are cross-brand threshold functions such that* (b, y) *is chosen if and only if:*

$$V_b \in [g_b(y), g_b(y+1)), \tag{3.5}$$

$$\forall k < b,\ V_k \leq h_{bk}(y),\ and \tag{3.6}$$

$$\forall k > b,\ V_k < h_{bk}(y), \tag{3.7}$$

where $g_b(0) \equiv -\infty$ *and* $g_b(\bar{y}_b + 1) \equiv \infty$*. The threshold function* $g_b(\cdot)$ *may depend on* x *and each function* $h_{bk}(\cdot) : k \neq b$*, may depend on both* v_b *and* x*.*

4 A Parametric Example: A Partially Ordered Probit Model

In this section we consider a simple parametric example with two firms $b \in \mathcal{B} = \{1, 2\}$, each selling a low-quality product offering $(Y = 1)$ and a high-quality product offering $(Y = 2)$, so that $\mathcal{Y}_1 = \mathcal{Y}_2 \equiv \{0, 1, 2\}$. We specify the utility functions $u(b, \cdot, \cdot, \cdot) : \mathcal{Y}_b \times \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ for each $b \in \mathcal{B}$ and $y \in \{1, 2\}$ as:

$$u(b, y, x, v_b) \equiv y \times (x_b\beta_b + v_b) - \alpha_{by} \tag{4.1}$$

where $\theta \equiv (\beta_1, \beta_2, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})'$ are the parameters of the utility function. The utility of choosing the outside option is normalized to zero, such that $u(b, 0, x, v_b) \equiv 0$.

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

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

where

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

denote threshold parameters. Some algebra reveals that

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

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,

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

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 $y = 1$.

With α_{b1}, α_{b2} fixed parameters that do not depend on observable variables, the inequality $\alpha_{b2} \leq 2\alpha_{b1}$ implies that $B = b$ and $Y = 1$ never occurs. So if $B = b$ and $Y = 1$ are indeed observed in the data, then it would be sensible to simply impose the inequality $\alpha_{b2} > 2\alpha_{b1}$, and parameter configurations with $\alpha_{b2} \leq 2\alpha_{b1}$ would imply a log-likelihood of $-\infty$.

We further restrict $V = (V_1, V_2)$ to be bivariate normally distributed with mean zero and variance

$$\Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}.$$

Given the parametric specification (4.1) for u, the resulting regions of unobserved variables $\mathcal{V}_{by}$ defined in (3.2) take the form of convex polytopes in $\mathbb{R}^2$. Figure 1 gives an example illustrating these regions for a particular parameter vector θ and a given value of the conditioning variables x in which the inequality $\alpha_{b2} > 2\alpha_{b1}$ on the left hand side of (4.4) holds.

The resulting choice model can alternatively be cast as a multinomial probit model with a 4 dimensional jointly normal unobservable, but with a singular variance matrix due to the ordered structure of the within brand choice. This is problematic, complicating standard arguments for identification and inference. See for example Weeks and Orme (1999) and Poirier and Kapadia (2012) for models in which similar issues arise when agents make mul-

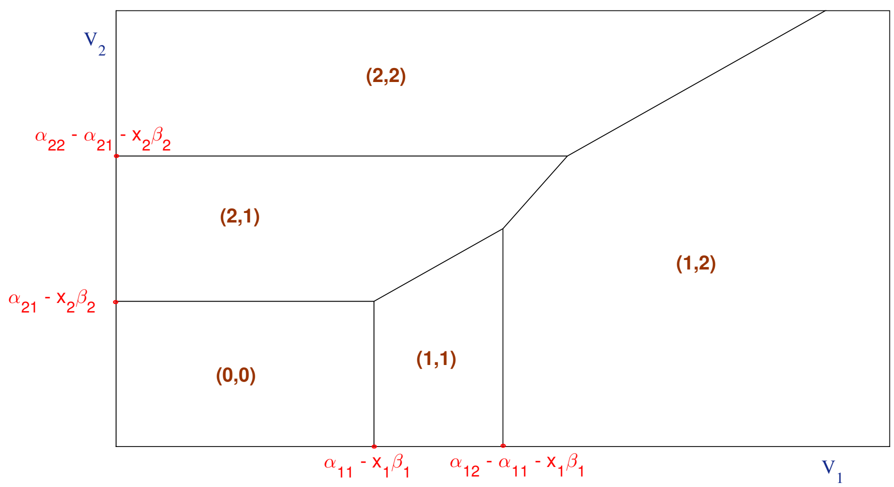

Figure 1: Regions of unobservables V resulting in each choice of $(b, y) \in \mathcal{M}_{BY}$ with utility as specified in (4.1).

tiple discrete choices simultaneously. In the next section we provide identification analysis by making explicit use of the particular structure of the partially ordered probit model.

The specification set out here treats the thresholds $(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$ as fixed parameters to be estimated. Fixed threshold specifications for ordered probit and logit models are common, and we begin our analysis with this restriction. It is however conceptually straightforward to allow these thresholds to be functions of observable parameters. This is important in our application, where observed prices may affect the utility of purchasing each product. We thus begin with identification analysis with the fixed threshold specification in Section 4.1, before considering the case where the thresholds can be a function of an observable variable in Section 4.2.

4.1 Fixed Thresholds

We now specialize the characterization of the set $\mathcal{S}_0(\mathcal{X})$ from Proposition 1 and the form of the conditional choice probabilities given in Theorem 1 to the case of the partially ordered probit model. In this section we have admissible structures $\mathcal{S} \equiv \mathcal{U} \times \mathcal{G}$, where

$$\mathcal{U} \equiv \left\{ \begin{array}{c} u : \mathcal{B} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{V} \rightarrow \mathbb{R} : \ u(b, y, x, v) \equiv y \times (x_b \beta_b + v_b) - \alpha_{by} \\ \text{for some } \theta \equiv \{(\beta_b, \alpha_{b1}, \alpha_{b2}) : b \in \mathcal{B}\} \in \Theta. \end{array} \right\}, \tag{4.6}$$

where Θ is a compact subset of Euclidean space, and

$$\mathcal{G} \equiv \left\{ \text{bivariate normal distribution functions } G \text{ with variance } \Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} \right\}. \quad (4.7)$$

Each admissible utility function and distribution pair (u, G) is completely specified given (θ, Σ), so we simply write (θ, Σ) to denote the corresponding structure $(u, G) \in \mathcal{S}$, henceforth writing $\mathcal{S}_0(\mathcal{X})$ as a set of parameterizations (θ, Σ) for structures that lie in the identified set.

Let $\mathbf{\Lambda}$ denote some set of positive definite matrices Σ with $\Sigma_{11} = 1$, and let $G(\cdot; \Sigma)$ be the distribution function for the bivariate normal distribution with zero mean and variance Σ of the form specified in (4.7). Define

$$p_{by}(x; \theta, \Sigma) \equiv G(\mathcal{V}_{by}(x; u); \Sigma),$$

to be the conditional probability that $B = b$ and $Y = y$ given $X = x$ generated by utility function u from (4.6) with parameter vector θ and distribution from (4.7) with parameter Σ. Then by definition the identified set as given in (3.1) is

$$\mathcal{S}_0(\mathcal{X}) \equiv \left\{ (\Sigma, \theta) \in \mathbf{\Lambda} \times \Theta : \forall (b, y) \in \mathcal{B} \times \mathcal{Y},\ p_{by}(x; \theta, \Sigma) = f_x^0(b, y) \text{ a.e. } x \in \mathcal{X} \right\}.$$

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

$$\mathcal{S}_0(\mathcal{X}) \equiv \underset{(\Sigma,\theta) \in \mathbf{\Lambda} \times \Theta}{\text{argmax}}\, E\left[\ln p_{BY}(X; \theta, \Sigma)\right].$$

Using the conditional choice probability integrals of Theorem 1 this becomes

$$\mathcal{S}_0(\mathcal{X}) = \underset{(\Sigma,\theta) \in \mathbf{\Lambda} \times \Theta}{\text{argmax}}\, \mathcal{L}(\theta, \Sigma),$$

with $\mathcal{L}(\theta, \Sigma)$ the expected log-likelihood:

$$\mathcal{L}(\theta, \Sigma) \equiv E\left[\ln p_{BY}(X; \theta, \Sigma)\right] = E_x E\left[\ln p_{BY}(X; \theta, \Sigma) \,|X = x\right].$$

Using the parametric structure set out above we have

$$E\left[\ln p_{BY}\left(X;\theta,\Sigma\right)|X=x\right]\equiv\sum_{(b,y)\in\mathcal{B}\times\mathcal{Y}}f_x^0\left(b,y\right)\ln\left(\int\limits_{g_b(y;x,\theta)}^{g_b(y+1;x,\theta)}\int\limits_{-\infty}^{h_{by}(y,x,v,\theta)}\phi_2\left(v,\Sigma\right)dv_{3-b}dv_b\right),$$

where $\phi_2\left(\cdot,\Sigma\right)$ denotes the density of a zero mean bivariate normal random variable with variance Σ and where

$$g_b\left(y;x,\theta\right)\equiv\alpha_{by}-\alpha_{b,y-1}-X_b\beta_b,\quad\alpha_{b0}\equiv0,$$

and for all $d\neq b$,

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

Thus each $p_{by}\left(x;\theta,\Sigma\right)$ takes the form of an integral over a region defined by inequalities that are linear in the parameters θ, equivalently

$$p_{by}\left(x;\theta,\Sigma\right)=\int\limits_{\mathbb{R}^2}\phi_2\left(v,\Sigma\right)1\left[v\in\mathcal{V}_{by}\left(x;\theta\right)\right]dv.$$

Written in this form it is straightforward to verify that $p_{by}\left(x;\theta,\Sigma\right)$ is log-concave for each value (b,y,x). This in turn implies that the maximizers of $\mathcal{L}\left(\theta,\Sigma\right)$ for any fixed Σ comprise a convex set.

Theorem 2 *Suppose that Restrictions A1-A7 hold, that $u\in\mathcal{U}$ defined in (4.6), and G is known with log-concave density g. Then the identified set for θ is*

$$\Theta^*\equiv\arg\max_{\theta\in\Theta}\mathcal{L}\left(\theta,G\right),$$

with the expected log-likelihood

$$\mathcal{L}\left(\theta,G\right)\equiv\sum_{(b,y)\in\mathcal{B}\times\mathcal{Y}}f_x^0\left(b,y\right)\ln\int\limits_{\mathbb{R}^{|\mathcal{B}|}}g\left(v\right)1\left[v\in\mathcal{V}_{by}\left(x;\theta\right)\right]dv,$$

concave in θ.

Many commonly used distributions are log-concave, with the normal distribution being 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 set delivered by Theorem 2 for each $G \in \mathcal{G}$. Under some additional but mild conditions on the variation in observable variables X, a known G in fact delivers point identification, as stated in Theorem 3 below.

The first part of Theorem 3, which establishes identification of $(\alpha_{b1}, \beta_b')'$ for each $b \in \{1, 2\}$, is a restatement of a result initially proven in Theorem 2 of Aradillas-Lopez and Rosen (2014), up to minor changes in notation. The second part then provides a straightforward extension applicable to prove identification of the additional parameters $(\alpha_{12}, \alpha_{22})$. The reason the result from Aradillas-Lopez and Rosen (2014) applies is the equivalence of the conditional probability of consumer choosing not to purchase, i.e. $p_{b0}(X; \theta, \Sigma)$ 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 (2014).[2] While both models feature the same conditional probabilities for these particular outcomes, the rest of their observable implications differ. The simultaneous equations model of Aradillas-Lopez and Rosen (2014) 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, enabling estimation of and inference on the resulting identified set by maximum likelihood.

The full result is now provided for completeness.

Theorem 3 *Suppose that Restrictions A1-A7 hold and that we have the probit structure* $\mathcal{S} = \mathcal{U} \times \mathcal{G}$ *given in (4.6) and (4.7) with singleton* $\mathcal{G}$ *so that* Σ *is known, with* $|\rho| < 1$ *and* $\sigma > 0$. *For each* $b \in \{1, 2\}$, *let* $Z_b \equiv (1, -X_b)$. *Then if (i) for each* $b \in \{1, 2\}$ *there exists no proper linear subspace of the support of* Z_b *that contains* Z_b *with probability one, and (ii) for all conformable column vectors* c_1, c_2 *with* $c_2 \neq 0$, *we have that either (a)* $\mathbb{P}\{Z_2 c_2 \leq 0 | Z_1 c_1 < 0\} > 0$; *or (b)* $\mathbb{P}\{Z_2 c_2 \geq 0 | Z_1 c_1 > 0\} > 0$, *then* θ *is point identified.*

The Theorem above shows that under conditions that guarantee sufficient variation in ex-

[2] Indeed, Theorem 2 of Aradillas-Lopez and Rosen (2014) can also be applied to simplify characterization of the identified set for a subset of parameters in simultaneous binary models studied in e.g. Heckman (1978), Bresnahan and Reiss (1991), and Tamer (2003) when the large support conditions stated in Tamer (2003) do not hold.

ogenous variables X, θ is point identified. The first of these, condition (i), is standard. Note that this requires that each X_b contains no constant components. Condition (ii) first appeared in Aradillas-Lopez and Rosen (2014). It restricts the joint distribution of Z_1 and Z_2, requiring that conditional on $Z_1 c_1$ negative (positive), $Z_2 c_2$ takes nonpositive (nonnegative) values with nonzero probability. Intuitively this condition helps to achieve identification because when there is a change in values of z such that the indices $z_1\left(\tilde{\delta}_1-\delta_1\right)$ and $z_2\left(\tilde{\delta}_2-\delta_2\right)$ move in the same direction, there is for fixed G a strict difference in the induced change in the conditional probability of choosing the outside alternative. This implies that the lower left rectangles labeled $(0,0)$ in Figure 1 associated with parameter vectors $\tilde{\delta}$ and δ are such that one is strictly contained in the other. This in turn implies different conditional probabilities for the outside option, so that $\tilde{\delta}$ and δ are not observationally equivalent. Note that this condition is automatically satisfied under a large support restriction on a component of either X_1 or X_2, for example if X_{11} has positive density on the real line conditional on any realization of X_2, with $\beta_{11}\neq 0$, but is considerably weaker and does not rely on an identification at infinity argument.

Theorem 3 requires that the distribution of unobserved heterogeneity G is known, which is a strong restriction. However, the Theorem has useful implications for settings where G is not known, but rather restricted to belong to some set of admissible distributions $\mathcal{G}$. Under the stated conditions, we have that for each $\tilde{G}\in\mathcal{G}$, if we were to assume $V\sim\tilde{G}$, that is $G=\tilde{G}$, there would be a singleton identified set with element denoted $\theta\left(\tilde{G}\right)$. Thus the identified set can only consist of parameter values for θ that are $\theta\left(\tilde{G}\right)$ *for some* $\tilde{G}\in\mathcal{G}$. Furthermore, since θ is identified under the restriction that $V\sim\tilde{G}$, $\theta\left(\tilde{G}\right)$ is consistently estimable via maximum likelihood. The following corollary formalizes these results, which are immediate consequences of Theorems 1 and 3.

Corollary 1 *Let all the restrictions of Theorem 3 hold except for the restriction that $\mathcal{G}$ is singleton. Then the identified set for θ, denoted $\mathcal{S}_0(\mathcal{X})$ is a subset of the set $\mathcal{S}_{\blacklozenge}(\mathcal{X})$ defined as*

$$\mathcal{S}_{\blacklozenge}(\mathcal{X})\equiv\left\{(\theta,\Sigma)\in\Theta\times\mathbf{\Lambda}:\theta=\theta^{*}(\Sigma)\right\},$$

where for each $\Sigma\in\mathbf{\Lambda}$, $\theta^{*}(\Sigma)$ is the unique solution to

$$\max_{\theta\in\Theta}\mathcal{L}(\theta,\Sigma).$$

Moreover, each $\theta^*(\Sigma)$ is consistently estimated by the maximum likelihood estimator

$$\hat{\theta}(\Sigma) \equiv \arg\max \frac{1}{n}\sum_{i=1}^{n}\left[\ln p_{b_i y_i}(x_i;\theta,\Sigma)\right].$$

Corollary 1 shows that the identified set for parameters θ in this model must maximize $\mathcal{L}(\theta,\Sigma)$ for some Σ. Put another way, the set $\mathcal{S}_{\blacklozenge}(\mathcal{X})$ is an outer region for the identified set $\mathcal{S}_0(\mathcal{X})$, in the sense that $\mathcal{S}_0(\mathcal{X}) \subseteq \mathcal{S}_{\blacklozenge}(\mathcal{X})$. In principle, an analogous implication can be extended to settings where $\mathcal{G}$ is nonparametrically specified, keeping the parametric structure for u of (4.6), by profiling over $G \in \mathcal{G}$ rather than $\Sigma \in \mathbf{\Lambda}$.

4.2 Variable Thresholds

As in classical ordered choice models, in some applications it may be desirable to allow the threshold parameters $(\alpha_{11},\alpha_{12},\alpha_{21},\alpha_{22})$ to depend functionally on observable variables. In our empirical application in Section 6 prices for each alternative are observed, and it is reasonable to allow the thresholds to depend on the menu of prices each consumer faces. The price menu faced by each consumer depends on the prices offered in the store in which they shop, which is observable in our data. We presume that all covariates and all prices are jointly independent of unobservable heterogeneity (V_1,V_2), a reasonable assumption if consumers are thought to choose the store where their purchase is made independently of the price menu for the product studied, here safety razors. If instead consumers are thought to choose their store based on the price of razors in that store, then prices could be correlated with unobserved heterogeneity, but this is assumed not to be the case here.

For each $b=1,2$ and $y=1,2$ parameters α_{by} are determined by

$$\alpha_{by} = \delta_b + g(p_{by},\gamma_b), \tag{4.8}$$

where $p_{by} \in \mathbb{R}$ denotes the price of alternative (b,y) and for each $b \in \mathcal{B}$, 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}. The utility function for choice (b,y) is now

$$u(b,y,X,v_b) \equiv y(x\beta_b + v_b) - \delta_b - g(p_{by},\gamma_b),$$

while the utility of the outside alternative continues to be normalized to zero. The within b optimal choice Y_b^* is still determined by (4.2) and (4.3), but with (4.8) as above.

Various specifications for $g\left(p_{by},\gamma_b\right)$ are possible, and in each specification the parameters $\delta_1,\delta_2,\gamma_1,\gamma_2$ pin down the trade-off between quality and price. We focus here on the linear specification

$$g\left(p_{by},\gamma_b\right)=\gamma_b p_{by}, \tag{4.9}$$

with $\gamma_b \geq 0$ such that utility is decreasing in price for each choice (b,y). Other possible specifications include the CRRA or isoelastic utility specification $g\left(p_{by},\gamma_b\right)=\left(1-\gamma_b\right)^{-1}p_{by}^{1-\gamma_b}$ with $\gamma_b>0$ or the exponential utility specification $g\left(p_{by},\gamma_b\right)=1-\exp\left(-\gamma_b p_{by}\right)$ for some $\gamma_b \in \mathbb{R}$.

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 empirically relevant. This is because unlike the brand-specific covariates X_b, prices vary across the vertical dimension y within brand. For the sake of explanation, let $Z=\left(X_1,X_2,P_{11},P_{12},P_{21},P_{22}\right)$. In contrast to the fixed threshold specification, it is possible now 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.

Under the linear price specification (4.9) used here, the functional form of the utility function for $Y=1$ can be manipulated so as to establish point identification of the utility function parameters $\left(\beta_b,\delta_b,\gamma_b\right)$, $b=1,2$, by application of Theorem 3. To see how, define

$$\mathcal{Z}^{*}\equiv\left\{z\in\operatorname{Supp}(Z):\mathbb{P}\left[(B,Y)=(b,1)\,|Z=z\right]>0,\text{ each } b=1,2\right\}.$$

Then for any $z\in\mathcal{Z}^{*}$,

$$\mathbb{P}\left[Y=0|Z=z\right]=\Phi_2\left(z_1\vartheta_1,z_2\vartheta_2;\Sigma\right)$$

where for each b, $\vartheta_b\equiv\left(\delta_b,\gamma_b,\beta_b\right)'$ and $Z_b\equiv\left(1,P_{b1},-X_b\right)$. Now the same approach used to guarantee point identification of parameters when Σ is known in the fixed threshold setting of Section 4.1 can be applied by comparing $\Phi_2\left(z_1\vartheta_1,z_2\vartheta_2;\Sigma\right)$ with $\Phi_2\left(z_1\tilde{\vartheta}_1,z_2\tilde{\vartheta}_2;\Sigma\right)$ for $\left(\vartheta_1,\vartheta_2\right)\neq\left(\tilde{\vartheta}_1,\tilde{\vartheta}_2\right)$ as in lines (A.2) and (A.3) in the proof of Theorem 3. The formal result is stated in the following Corollary.

Corollary 2 *Let the same restrictions hold as in Theorem 3. For each* $b \in \{1,2\}$*, let* $Z_b \equiv (1, P_{b1}, -X_b)$*. Then if (i) for each* $b \in \{1,2\}$ *there exists no proper linear subspace of the support of* Z_b *that contains* Z_b *with probability one conditional on* $Z \in \mathcal{Z}^*$*, and (ii) for all conformable column vectors* c_1, c_2 *with* $c_2 \neq 0$*, we have that either (i)* $\mathbb{P}\{Z_2 c_2 \leq 0 | Z_1 c_1 < 0, Z \in \mathcal{Z}^*\} > 0$*; or (ii)* $\mathbb{P}\{Z_2 c_2 \geq 0 | Z_1 c_1 > 0, Z \in \mathcal{Z}^*\} > 0$*, then* θ *is point identified.*

The Corollary establishes conditions whereby model parameters are point-identified if the distribution of unobservable heterogeneity is known with variable thresholds satisfying a linear specification. In practice, it may not be desirable to restrict the distribution of unobservable heterogeneity to be known. This will not be imposed in the next two sections, and we consequently allow for the possibility that the identified set is not a singleton. Trivially, by similar reasoning a result analogous to Corollary 1 also holds.

Point identification of utility parameters (up to scale) for either brand b may alternatively be achieved by imposing a large support restriction on unobservable heterogeneity. This would require that prices have full support on $\mathbb{R}^2_+$ for the other brand, in addition to mild rank conditions. Intuitively, the probability of $B = b$ conditional on $Z = z$ could then be made arbitrarily close to one by considering z with arbitrarily large values for the prices of the other brand's product offering. Conditional on such values of z, the conditional probabilities become arbitrarily close to that of a simple ordered probit model from brand b's product offerings, so that the usual rank condition establishes point identification. In practice however it is hard to argue that the price of any good has full support on $[0, \infty)$, so we do not consider this condition any further.

5 Likelihood Computation and Inference

In this section ζ will be used to denote the full vector of model parameters of the bivariate probit model with variable threshold specification given by (4.8) and (4.9). 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. 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 $k = 1, ..., \dim(\zeta)$.

For inference we use Procedure 3 of CCT to construct confidence intervals for each individual element of ζ. The approach does not require point identification. It is designed to perform inference on individual parameter components in models in which the identified set can be represented as the set of maximizers of a likelihood or by a system of moment equalities and inequalities. We chose their third procedure for its combination of ease of implementation and good performance reported in CCT. The approach entails collecting the set of values for the parameter component such that a profile likelihood ratio statistic is no greater than the corresponding quantile of a χ_1^2 random variable. When there is point identification, confidence intervals constructed this way have exact asymptotic coverage. As CCT show, the approach can be conservative when there is partial identification, but only to a limited extent when the nominal level of the confidence sets considered is roughly 0.85 or greater.

The choice probabilities implied by the partially ordered probit model – and which must be computed in order to compute the likelihood – are of the form set out in (3.4). These choice probabilities must be computed in order to compute the log-likelihood at candidate parameter values ζ for each observed value of conditioning variables z_i. In our application in Section 6 there are two brands, so (3.4) takes the form of a bivariate integral. The choice probabilities can thus be computed using numerical integration or by way of 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, likely due to the nonlinear nature of the objective function. With (ρ, σ) unknown the log-likelihood is generally not concave in parameters, and while it is continuous, it is not everywhere differentiable due to points at which individuals are indifferent between choosing the best option among the competing brands.

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. Instead, the alternative formulation of the choice probabilities involves univariate and bivariate normal CDFs evaluated at functions of parameters and observable variables. Software was used that vectorizes application of these CDFs, performing fast evaluation of the CDFs at each component of a vector of values in one function call.[3] This enabled computing the likelihood contribution for each observation in the data through use of the vectorized function, rather than performing numerical integration or computing simulated probabilities separately for each observation.

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

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 5.2 then explains how Procedure 3 of CCT was implemented and examines the performance of the approach in Monte Carlo experiments.

5.1 Computation of Choice Probabilities

In the partially ordered probit model exposited in Section 4, application of (3.4) gives the following representation for the conditional choice probabilities:

$$\forall (b,y) \in \overline{\mathcal{M}}_{BY}, \qquad p_{by}(x,\zeta) = \int_{\lambda_{b,y}}^{\lambda_{b,y+1}} \int_{-\infty}^{h_{by}(x,v,\theta)} \phi_2(v,\Sigma)\, dv_d dv_b, \tag{5.1}$$

$$p_0(x,\zeta) = \Phi_2(\alpha_{11} - x_1\beta_1, \alpha_{21} - x_2\beta_2; \Sigma), \tag{5.2}$$

where as before

$$h_{by}(x,v,\theta) \equiv \min_{\tilde{y}\in\{1,\dots,\bar{y}_d\}} \frac{1}{\tilde{y}} \left[y(x_b\beta_b + v_b) - (\alpha_{by} - \alpha_{d\tilde{y}})\right] - x_d\beta_d, \tag{5.3}$$

and $d \equiv 3 - b$ denotes the brand other than b. The inner integral in (5.1) can then be replaced by decomposing the joint density of V_b and V_d with the product of the marginal density of V_b with the conditional density of V_d given V_b and integrating with respect to v_d. This gives

$$p_{by}(x,\zeta) = \frac{1}{\sigma_b} \int_{\lambda_{b,y}}^{\lambda_{b,y+1}} \Phi\left(\frac{h_{by}(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi\left(\frac{z}{\sigma_b}\right) dz. \tag{5.4}$$

To remove the need to simulate or numerically approximate the above integral, conditional choice probabilities $p_{by}(x,\theta)$ 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*, the utility specification*

$$u(b,y,x,v_b) \equiv \begin{cases} y \times (x_b\beta_b + v_b) - \alpha_{by}, & \text{if } y \in \{1,2\} \\ 0, & \text{if } y = 0 \end{cases}$$

as in (4.1) and with $V = (V_1, V_2)$ 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 $\sigma > 0$ as specified in (4.7). Then the conditional choice probabilities for each $b = 1, 2$ and $y = 1, 2$ can be expressed as

$$p_{by}(x, \zeta) = \begin{pmatrix} 1\left[z_{by}^* < \lambda_{b,y+1}\right] \Delta\left(\sigma_b^{-1} \max\left\{z_{by}^*, \lambda_{by}\right\}, \sigma_b^{-1}\lambda_{b,y+1}, m_1^+, m_2^+\right) \\ +1\left[z_{by}^* > \lambda_{b,y}\right] \Delta\left(\sigma_b^{-1}\lambda_{by}, \sigma_b^{-1} \min\left\{z_{by}^*, \lambda_{b,y+1}\right\}, m_1^-, m_2^-\right) \end{pmatrix}, \tag{5.5}$$

where λ_{b1} and λ_{b2} are as defined in (4.3), for any reals h, k, c_1, c_2,

$$\Delta(h, k, m_1, m_2) \equiv \Phi_2(k, m_1; m_2) - \Phi_2(h, m_1; m_2), \tag{5.6}$$

where $\Phi_2(a, b, \rho)$ denotes the probability that a bivariate normal random vector Z with mean zero and unit variance components with correlation ρ satisfies both $Z_1 \leq a$ and $Z_2 \leq b$, and for $d \equiv 3 - b$,

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

and

$$m_1^+ \equiv \frac{yx_b\beta_b + \alpha_{d2} - \alpha_{by} - 2x_d\beta_d}{\sqrt{\sigma_b^2 y^2 - 4\rho\sigma_b\sigma_d y + 4\sigma_d^2}}, \quad m_2^+ \equiv \frac{2\rho\sigma_d - \sigma_b y}{\sqrt{\sigma_b^2 y^2 - 4\rho\sigma_b\sigma_d y + 4\sigma_d^2}}, \tag{5.7}$$

$$m_1^- \equiv \frac{yx_b\beta_b + \alpha_{d1} - \alpha_{by} - x_d\beta_d}{\sqrt{\sigma_b^2 y^2 - 2\rho\sigma_b\sigma_d y + \sigma_d^2}}, \quad m_2^- \equiv \frac{\rho\sigma_d - \sigma_b y}{\sqrt{\sigma_b^2 y^2 - 2\rho\sigma_b\sigma_d y + \sigma_d^2}}. \tag{5.8}$$

5.2 Computation of Confidence Sets and Monte Carlo Experiments

Before applying the partially ordered probit model of Section 4 to study consumer preferences for razors in the U.K. market, we first conducted Monte Carlo experiments to investigate the finite sample performance of the inference procedure used.[4] 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 variable

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

threshold linear-in-price specification described by (4.8) and (4.9) was used.

To simulate data population parameter values were set as follows.

$$\gamma_1 = 1, \quad \gamma_2 = 0.8, \quad \delta_1 = -1.5, \quad \delta_2 = -1.2, \quad \rho = 0.5, \quad \sigma = 1, \tag{5.9}$$
$$\beta_1 = (1.3, 0.3, -0.1, -0.3, 0.7)', \quad \beta_2 = (1.0, 0.3, -0.1, -0.3, 0.7)'.$$

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 from a population distribution in which $\Pr[X_1 = 1] = 0.426$ and $\Pr[X_2 = 1] = 0.234$. 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. In the Monte Carlos these were generated from the Bernoulli distribution with parameters 0.4, 0.85, and 0.554, respectively. All components of X were generated independently of each other.

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 can be achieved. This is in fact 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, 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 (5.9). The expression (5.5) 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 investigation, choice probabilities computed using (5.5) conditional on several values of observable variables were compared to those obtained using the integral formula (5.4) and those obtained by simulation, and these were all found to be in close agreement up to negligible computation difference.

In order to perform inference on structural parameters the third procedure proposed by CCT was used. This is a particularly attractive approach for constructing confidence intervals for parameter components because it is obtained by inverting a likelihood ratio test statistic using a simple chi-square critical value[5]. Specifically, an asymptotic α-level confidence set for any individual parameter component, say ζ_k is

$$\widehat{M}^{\chi}_{\alpha,k} = \left\{ \mu \in \Upsilon_k : \inf_{\zeta \in \Upsilon : \zeta_k = \mu} Q_n(\zeta) \leq \chi^2_{1,\alpha} \right\}, \tag{5.10}$$

where $\chi^2_{1,\alpha}$ denotes the α quantile of the χ^2_1 distribution, and

$$Q_n(\zeta) \equiv 2n\left[L^*_n - L_n(\zeta)\right]$$

is the quasi likelihood ratio statistic, with $L^*_n \equiv \max_{\zeta \in \Upsilon} L_n(\zeta)$ and

$$L_n(\zeta) \equiv \sum_{i=1}^{n} \left[\ln p_{b_i y_i}(x_i; \zeta)\right]$$

[5]See CCT for sufficient conditions for their procedure to provide asymptotically valid confidence intervals for the *identified set* of parameter components ζ_k, each k. For coverage of the true parameter value ζ_k itself, it may be possible to establish either weaker sufficient conditions or strictly smaller confidence intervals – questions we leave open to future research.

denoting the log-likelihood. Define the profile log-likelihood for ζ_k evaluated at any parameter value $\mu \in \Upsilon_k$ as

$$PL_{k,n}(\mu) \equiv \sup_{\zeta \in \Upsilon : \zeta_k = \mu} L_n(\zeta), \tag{5.11}$$

and the profile log-likelihood ratio statistic as

$$Q_{k,n}(\mu) \equiv \inf_{\zeta \in \Upsilon : \zeta_k = \mu} Q_n(\zeta) = 2n\left[L_n^* - PL_{k,n}(\mu)\right]. \tag{5.12}$$

In order to compare the empirical coverage frequencies of classical ML confidence intervals to those of the identification robust confidence intervals $\widehat{M}_{\alpha,k}^{\chi}$ 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.[6] The optimization routine also returned a numerical approximation to the Hessian at the optimal value, and this was used to construct standard errors for 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. Classical maximum likelihood confidence intervals for each component of ζ were thus computed for the sake of comparison to $\widehat{M}_{\alpha}^{\chi}$, as also suggested by CCT.

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, confidence intervals constructed using this ML estimator and the classical asymptotic normal approximation should be expected to perform well, and in our Monte Carlo experiments this was indeed the case. We refer to this as the "oracle ML" procedure in the results reported below, whereas the maximum likelihood procedure treating ρ and σ as additional parameters to estimate is referred to as "feasible ML".

The steps described so far are sufficient to construct oracle ML and feasible ML con-

[6]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.

DGP 1	Realized Coverage Percent, $n = 200, 500, 1000, 2000$		
Parameter	**Oracle ML**	**Feasible ML**	**Profile LRTS**
γ_1	99, 97, 98, 95	97, 95, 96, 96	96, 96, 96, 95
γ_2	95, 96, 96, 92	95, 95, 96, 99	97, 96, 96, 99
δ_1	94, 94, 93, 96	96, 91, 94, 94	97, 93, 95, 93
δ_2	97, 94, 97, 94	95, 95, 93, 96	94, 94, 95, 96
β_{11}	95, 97, 95, 95	94, 98, 98, 94	91, 98, 97, 95
β_{12}	96, 98, 96, 94	96, 98, 94, 95	95, 97, 93, 95
β_{13}	96, 97, 96, 95	95, 97, 97, 98	93, 97, 96, 96
β_{14}	95, 96, 99, 93	95, 97, 97, 95	93, 96, 97, 95
β_{15}	94, 98, 97, 96	94, 92, 96, 98	93, 92, 96, 97
β_{21}	96, 97, 98, 95	93, 96, 95, 97	94, 97, 96, 97
β_{22}	96, 96, 94, 97	97, 93, 96, 98	97, 95, 96, 97
β_{23}	98, 96, 94, 96	98, 97, 92, 97	94, 96, 91, 97
β_{24}	98, 93, 96, 99	96, 93, 94, 95	92, 94, 94, 94
β_{25}	94, 96, 96, 96	90, 92, 93, 99	92, 93, 93, 98
ρ	–	89, 93, 96, 96	94, 93, 95, 96
σ	–	91, 96, 94, 96	90, 96, 94, 95

Table 1: Monte Carlo coverage frequencies out of 100 simulations for sample sizes $n = 200, 500, 1000, 2000$ for DGP1, as described in the text.

fidence intervals for each ζ_k. This is done by taking each of the likelihood estimators for ζ_k and adding and subtracting 1.96 times their respective standard errors. This makes it easy to compute the empirical frequency with which these confidence intervals contain the population ζ_k in Monte Carlo experiments. In order to compute Monte Carlo empirical coverage frequencies of $\widehat{M}^{\chi}_{\alpha,k}$ for ζ_k in simulations, one also needs to compute $PL_{k,n}(\zeta_k)$ and then check whether $Q_{k,n}(\mu) \leq \chi^2_{1,\alpha}$ in each simulation. A profile likelihood function was computed, employing the solnp function from the package Ghalanos and Stefan (2015) to compute $\sup_{\zeta \in \Upsilon : \zeta_k = \mu} L_n(\zeta)$.[7]

The empirical coverage frequency of the three different procedures for DGPs 1-3 out of 100 Monte Carlo repetitions for each sample size are reported in Tables 1, 2, and 3. The target coverage level in each case was 0.95. A first observation is that the Oracle ML procedure that makes use of knowledge of ρ and σ – which are not known in practice – does quite well across DGPs and sample sizes. This is not surprising. There are some cases

[7] In the constrained optimization 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 2**</u>	**Realized Coverage Percent,** $n = 200, 500, 1000, 2000$		
Parameter	**Oracle ML**	**Feasible ML**	**Profile LRTS**
γ_1	$93, 96, 91, 95$	$95, 96, 91, 96$	$93, 95, 92, 96$
γ_2	$94, 94, 96, 98$	$88, 88, 93, 95$	$94, 89, 95, 95$
δ_1	$93, 93, 95, 97$	$95, 93, 95, 97$	$90, 93, 95, 97$
δ_2	$97, 93, 98, 95$	$88, 93, 95, 93$	$91, 91, 92, 92$
β_{11}	$99, 95, 92, 95$	$93, 95, 92, 94$	$92, 95, 95, 93$
β_{12}	$91, 97, 94, 96$	$94, 97, 93, 96$	$93, 96, 93, 96$
β_{13}	$91, 94, 93, 97$	$92, 95, 95, 97$	$91, 95, 95, 97$
β_{14}	$98, 96, 97, 96$	$98, 96, 94, 95$	$97, 95, 94, 96$
β_{15}	$93, 93, 91, 94$	$90, 91, 93, 91$	$93, 93, 94, 93$
β_{21}	$95, 95, 97, 94$	$89, 93, 92, 92$	$92, 95, 93, 93$
β_{22}	$91, 96, 95, 96$	$83, 93, 91, 98$	$92, 95, 90, 95$
β_{23}	$96, 95, 96, 96$	$95, 98, 96, 98$	$98, 97, 93, 98$
β_{24}	$95, 96, 96, 98$	$89, 95, 91, 96$	$90, 96, 94, 96$
β_{25}	$95, 95, 94, 95$	$88, 94, 92, 98$	$91, 95, 93, 98$
ρ	–	$77, 83, 82, 95$	$84, 96, 93, 96$
σ	–	$89, 93, 94, 95$	$95, 92, 93, 93$

Table 2: Monte Carlo coverage frequencies out of 100 simulations for sample sizes $n = 200, 500, 1000, 2000$ for DGP2, as described in the text.

<u>**DGP 3**</u>	**Realized Coverage Percent,** $n = 200, 500, 1000, 2000$		
Parameter	**Oracle ML**	**Feasible ML**	**Profile LRTS**
γ_1	$96, 89, 90, 95$	$97, 94, 93, 98$	$97, 93, 93, 97$
γ_2	$93, 98, 96, 92$	$86, 92, 93, 97$	$91, 91, 92, 96$
δ_1	$96, 94, 96, 98$	$96, 93, 97, 98$	$95, 94, 98, 98$
δ_2	$93, 97, 96, 90$	$83, 89, 92, 93$	$90, 92, 93, 94$
β_{11}	$98, 94, 92, 98$	$96, 96, 92, 97$	$96, 95, 91, 97$
β_{12}	$93, 97, 90, 96$	$95, 98, 92, 95$	$92, 97, 92, 95$
β_{13}	$93, 91, 94, 97$	$95, 94, 95, 96$	$93, 93, 95, 96$
β_{14}	$99, 95, 95, 95$	$99, 95, 96, 94$	$98, 95, 96, 93$
β_{15}	$96, 96, 97, 96$	$92, 96, 95, 93$	$92, 97, 95, 92$
β_{21}	$93, 96, 96, 90$	$91, 94, 92, 94$	$96, 97, 92, 94$
β_{22}	$95, 95, 95, 96$	$97, 95, 95, 96$	$96, 94, 95, 96$
β_{23}	$96, 96, 95, 96$	$99, 97, 95, 99$	$98, 96, 94, 98$
β_{24}	$94, 95, 92, 96$	$93, 98, 93, 98$	$94, 95, 95, 97$
β_{25}	$93, 96, 94, 97$	$90, 90, 89, 95$	$93, 94, 91, 96$
ρ	–	$78, 90, 92, 93$	$89, 93, 95, 92$
σ	–	$85, 92, 92, 95$	$92, 91, 91, 93$

Table 3: Monte Carlo coverage frequencies out of 100 simulations for sample sizes $n = 200, 500, 1000, 2000$ for DGP3, as described in the text.

in which the observed coverage frequencies are below 0.95, but with 14 parameters, three DGPs, and four sample sizes comprising a total of 168 different reported empirical coverage probabilities, some variation should be expected. The lowest coverage probability (which occurred just once) was 0.89. As should be expected, feasible ML did not perform quite as well, but it performed reasonably well at least in some cases. A second observation is that the profile LRTS test, robust to a potential lack of point identification, also generally performs well. Further, unlike the oracle ML procedure, it can be used in practice. A third observation is that the feasible ML procedure does not always perform as well as the other two. For DGP1 its performance is fine, but for DGP2 in some cases it substantially under-covers the true parameter value. The degree of undercoverage seems to fall at larger sample sizes, but even at $n = 1000$ the empirical coverage for ρ is only 0.82.[8] On the other hand the LRTS procedure does not exhibit nearly as severe degrees of undercoverage. In general the feasible ML and LRTS procedures applied to DGP2 produce empirical coverage probabilities that are quite close, i.e. within just two or three percent, except for the cases in which feasible ML severely undercovers, in which case the LRTS procedure performs considerably better with coverage closer to the nominal level than feasible ML. A similar observation holds up for DGP3. It may very well be that parameters ρ and σ are not identified in all DGPs, but we cannot say for sure.

In any case, CCT's profile LRTS procedure performs adequately, and seemingly better than confidence intervals using the classical asymptotically normal approximation of the feasible ML procedure for two of the three DGPs considered. Moreover, as CCT point out, their LRTS procedure is in fact asymptotically exact in point-identified regular models.

In the application carried out in the next section, our goal is not to compute an empirical coverage frequency (which is indeed unknown since ζ is unknown), but rather to construct confidence intervals by inverting the LRTS test to compute $\widehat{M}^{\chi}_{\alpha,k}$ as in (5.10). To do this we first computed L_n^* by maximizing the likelihood over all parameters, as described above. Then the following steps were additionally carried out for each $k = 1, ..., \dim(\zeta)$. First, the profile log-likelihood $PL_{k,n}(\mu)$ was computed on values of μ over a grid of values $\mathcal{M}$. The

[8]It should be noted that in a handful of cases with $n = 200$ – specifically two, five, and two cases for DGP1, DGP2, and DGP3, respectively – the Hessian produced by feasible ML was singluar. In these cases standard errors were treated as infinite, so that feasible ML failed to reject the true parameter in all such cases. That is, the reported results count these cases as covering the true parameter value, and even so empirical coverage frequencies for many parameters are well below the nominal level. The oracle ML Hessian was always positive definite, as was the feasible ML Hessian for sample sizes greater than 200.

values

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

were recorded. 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 $Q_{k,n}(\mu) \leq \chi^2_{1,\alpha}$ required for $\mu \in \widehat{M}^{\chi}_{\alpha,k}$. 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 $\widehat{M}^{\chi}_{\alpha,k}$ 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 $PL_{k,n}(\widetilde{\mu})$.
2. If $Q_{k,n}(\widetilde{\mu}) \leq \chi^2_{1,\alpha}$ 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 $\widehat{M}^{\chi}_{\alpha,k}$ 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 $\widehat{M}^{\chi}_{\alpha,k}$.

6 Application to Razor Blade Purchases

This section presents an application of the parametric model in Section 4 to the market for women's razor blades using consumer data from the United Kingdom in the early 2000s. We use household purchase data from the Kantar Worldpanel. The data comprise repeated observations of purchases made by a representative sample of U.K. households, obtained by the use of a handheld scanner used to record all households 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 in the UK, Gillette and Wilkinson Sword.

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

In our application, we consider households in which the primary shopper is a female between the age of 18 and 50 years old, who was observed purchasing either blades for a reusable non-electric women's razor (which we refer to as "system blades") or disposable women's razors. We focus on purchases of Gillette and Wilkinson Sword blades made in the time period $2004 - 2005$.[10] In each period it is assumed that consumers buy the brand-quality combination that maximizes their utility. The outside alternative consists of all those individuals observed buying a disposable razor blade in the period $2004 - 2005$. The total sample size consists of 4842 observations. Table 4 shows the observed market shares of Gillette and Wilkinson Sword system blades and disposable razors, while Table 5 shows the observed market shares for the years 2004-2005, conditional on buying either double or triple blade cartridges from either brand.

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

Table 4: Market shares for Gillette blades, Wilkinson Sword blades and disposable razors in 2004-2005.

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

The covariates used for each household are indicator variables for age of the shopper being between each of 18-30, 31-40 and 41-50, indicator variables for marital status, employment and the presence of more than one female in the household. Table 6 provides descriptive statistics.

For each observation in the sample the brand-quality combination of blades purchased is observed, in addition to the individual characteristics. For estimation we also make use of the prices of the razor blade cartridges available to consumers. For the razor blade purchased,

[10]Consumers who switch to a new refillable razor will not be present in the sample at the time of the switch, but they will appear the next time they buy cartridges for their handle.

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 6: Consumer Characteristics for 2004-2005.

we observe the total expenditure, w, as well as the pack size, v, and the number of packs, n, of the purchased razor blades in a specific store, in a specific month for a specific brand and quality level, in a specific pack size. The average per cartridge price of the purchased razor blades, in a specific month and store, and for the specific brand, blade type and pack size, is thus calculated by

$$p_{by} = \frac{w_{by}}{n_{by} v_{by}}. \tag{6.1}$$

We do not however directly observe the counterfactual price per cartridge faced by each consumer for the razor blade types that they did not actually purchase. The counterfactual prices were thus imputed using data on all cartridge purchases in $2004 - 2005$. To do this we estimated counterfactual prices p^c_{by} by using a best linear predictor of p_{by} under two different specifications:

$$p^c_{bymsv} = \beta_0 + 1(B=2)\beta_2 + 1(Y=3)\beta_3 + \sum_{m=1}^{24} 1(M=m)\beta_m + \sum_{s=1}^{15} 1(S=s)\beta_s + \sum_{v=1}^{3} 1(V=v)\beta_v + \varepsilon_{bymsv}, \tag{6.2}$$

$$p^c_{byms} = \beta_0 + 1(B=2)\beta_2 + 1(Y=3)\beta_3 + \sum_{m=1}^{24} 1(M=m)\beta_m + \sum_{s=1}^{15} 1(S=s)\beta_s + \varepsilon_{byms}, \tag{6.3}$$

where $M = \mathit{month}$, $S = \mathit{store}$, $B = 2$ corresponds to Wilkinson Sword, $Y = 3$ corresponds to triple blade, and $V = \mathit{pack\ size}$, and the intercept corresponds to the price of a Gillette double blade cartridge. The best linear predictors were computed using all purchases of cartridges in our data from the 24 months spanning $2004 - 2005$, 15 stores, and three different pack sizes. To impute counterfactual prices, the best linear predictor p^c_{bymsv} was then matched to each consumer according the actual month, store and (in the case of specification (6.2)) pack size purchased. Specification (6.2) was estimated using the average price per cartridge in a fixed month, in a specific store, for a specific brand-quality combination in a fixed pack size, while specification (6.3) estimates the counterfactual prices without conditioning on the pack size. We chose to differentiate between the two specifications as not all the blade-types and/or brands offer all pack sizes. For example, the double blade razor was only offered in a five cartridge pack. In order to deal with this, pack size was categorized according to small, medium and large when specification (6.2) was used, see Appendix B for further details.

Tables 7 and 8 report point estimates obtained by maximum likelihood, conventional maximum likelihood confidence intervals, and confidence intervals constructed as described in Section 5.2 following CCT using specifications (6.2) and (6.3) for counterfactual prices. Here by conventional confidence intervals, we mean that 1.96 standard errors obtained by inverting the Hessian form of the asymptotic variance are added and subtracted from the ML point estimate. When reported, the two different types of confidence intervals are found to be quite close to each other, suggesting either the possibility of point identification, or that the identified set is quite small. Note that the parameter vector point estimator reported here is a maximizer of the log-likelihood. Given the dimension of the parameter space it cannot be guaranteed with certainty that it is unique, and in any case this would not guarantee that the population expected log-likelihood has a unique maximizer. Nonetheless, from Redner (1981) we know that a point estimator defined as a maximizer of the log-likelihood will be contained in the identified set with probability approaching one as $n \to \infty$.

The estimates and confidence intervals in Table 7 lead to several observations. The coefficients γ_1 and γ_2 on the price charged for blades of both brands are positive, so that utility is measured to be descreasing in price, although the coefficient on price for Gillette cartidges (γ_1) is considerably smaller than the coefficient for Wilkinson Sword cartridges (γ_2), even after scaling by the estimate of the standard deviation (σ_2) of the unobservable component of utility from a Wilkinson Sword purchase. 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, as compared to the outside alternative. Likewise, the coefficient on the more females indicator for either brand is found to be negative. Coefficients 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 preference for quality as reflected by blades per cartridge across the two different brands.[11] In this case the Hessian of the log-likelihood computed at the maximizing parameter vector was found to be singular. The value of $\hat{\rho}$ indicates that the point estimate is on the boundary of the parameter space. Thus,

[11] Maximum likelihood produced an estimate for the correlation coefficient of 0.99999992, effectively indistinguishable from 1.

conventional maximum likelihood confidence intervals are not reported for this specification. However, this does not preclude computation of CCT confidence intervals by inverting the likelihood ratio statistic. The CCT confidence interval for ρ is very tightly concentrated around 1.

Table 8 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, 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 (6.3) for counterfactual prices. Moreover, the Hessian was nonsingular, and consequently conventional maximum likelihood confidence intervals for each parameter are reported alongside the CCT confidence intervals.[12] 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. 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 on 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 larger than it was using the previous specification, but of similar magnitude.

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

Parameter	ML Point Estimate	CCT CI	
γ_1	0.0617	(-0.0380, 0.1737)	
γ_2	4.2788	(3.4006, 5.6788)	
δ_1	-0.202	(-0.2485, -0.2020)	
δ_2	-4.0525	(-5.3889, -3.3035)	
$\beta_{1Age31-40}$	-0.1566	(-0.2358, -0.0772)	
$\beta_{1Age41-50}$	-0.2264	(-0.3202, -0.1323)	
$\beta_{1Married}$	-0.0444	(-0.1092, 0.0202)	
$\beta_{1Employed}$	-0.0382	(-0.1072, 0.0320)	
$\beta_{1Females}$	-0.2236	(-0.2920, -0.1559)	
$\beta_{2Age31-40}$	-0.2392	(-0.4837, 0.0042)	
$\beta_{2Age41-50}$	-0.3644	(-0.6418, -0.0868)	
$\beta_{2Married}$	-0.1864	(-0.3953, 0.0128)	
$\beta_{2Employed}$	-0.1157	(-0.3316, 0.0915)	
$\beta_{2Females}$	-0.6098	(-0.8768, -0.4080)	
ρ	1.0000	(0.9994, 1.0000)	
σ_2	2.6941	(2.3474, 3.3302)	

Table 7: 95% confidence intervals with prices as specified in (6.2).

Parameter	ML Point Estimate	Conventional CI	CCT CI
γ_1	-0.0643	(-0.1274, -0.0013)	(-0.1157, -0.0036)
γ_2	2.6258	(2.3479, 2.9037)	(2.3449, 2.9131)
δ_1	-0.1650	(-0.1862, -0.1438)	(-0.1850, -0.1467)
δ_2	-2.5891	(-2.7832, -2.3951)	(-2.7805, -2.3872)
$\beta_{1Age31-40}$	-0.1595	(-0.2378, -0.0813)	(-0.2369, -0.0822)
$\beta_{1Age41-50}$	-0.2162	(-0.3079, -0.1246)	(-0.3087, -0.1240)
$\beta_{1Married}$	-0.0814	(-0.1435, -0.0194)	(-0.1435, -0.01923)
$\beta_{1Employed}$	-0.0585	(-0.1321, 0.0152)	(-0.1266, 0.0098)
$\beta_{1Females}$	-0.2378	(-0.3085, -0.1670)	(-0.3045,-0.1710)
$\beta_{2Age31-40}$	-0.3057	(-0.4779, -0.1335)	(-0.4763, -0.1361)
$\beta_{2Age41-50}$	-0.4181	(-0.6207, -0.2154)	(-0.6209, -0.2171)
$\beta_{2Married}$	-0.1937	(-0.3342, -0.0532)	(-0.3319 ,-0.0553)
$\beta_{2Employed}$	-0.1126	(-0.2741, 0.0488)	(-0.262 ,0.0369)
$\beta_{2Females}$	-0.5249	(-0.6796, -0.3702)	(-0.6719, -0.3800)
ρ	0.9998	(0.9996, 1.0000)	(0.9996, 1.0000)
σ_2	2.1580	(2.0505, 2.2655)	(2.0334, 2.2734)

Table 8: 95% confidence intervals with prices as specified in (6.3).

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 a general characterization of the identified set of structures admitted by the model under mild shape restrictions. We further showed that with some additional restrictions, such as the parametric restrictions in Section 4, characterization of the identified set can be further simplified. General conditions under which such sets reduce to a singleton set are not easily obtained, but inference methods robust to the possibility of set identification can be used. This was demonstrated using a recently developed method for inference by Chen, Christensen, and Tamer (2018), which was found to perform well in Monte Carlo simulations. An empirical illustration was provided using data on razor blade cartridge purchases in the United Kingdom, a setting that features two dominant competing firms with vertically differentiated products.

References

Aradillas-Lopez, A., and A. M. Rosen (2014): "Inference in Ordered Response Games with Complete Information," CEMMAP working paper CWP36/14.

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

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.

Bresnahan, T. F., and P. J. Reiss (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48(1-2), 57–81.

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.

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).*

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.

Heckman, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–959.

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.

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

Poirier, D. J., and D. Kapadia (2012): "Multivariate Versus Multinomial Probit: When are Binary Decisions Made Separately also Jointly Optimal?," Working paper, University of California, Irvine.

Ponomareva, M., and E. Tamer (2011): "Misspecification in moment inequality models: back to moment equalities?," *Econometrics Journal*, 14(2), 186–203.

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.

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.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Models with Multiple Equilibria," *Review of Economic Studies*, 70(1), 147–167.

Weeks, M., and C. Orme (1999): "The statistical relationship between bivariate and multinomial choice models," Cambridge Working Papers in Economics wp9912.

A Proofs

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

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

for all $(u, G) \in \mathcal{S}$. Thus $\mathcal{S}_0(\mathcal{X})$ is contained in the set of maximizers of $Q(u, G)$. Consider now $\left(\tilde{u}, \tilde{G}\right) \notin \mathcal{S}_0(\mathcal{X})$. Then for some $(b, y) \in \mathcal{Y} \times \mathcal{B}$ there exists a positive measure set $\mathcal{X}^*\left(\tilde{u}, \tilde{G}\right)$ as defined in (3.3) on which $\tilde{G}\left(\mathcal{V}_{by}(x;\tilde{u})\right) \neq G_0\left(\mathcal{V}_{BY}(x;u_0)\right) = f_x^0(b, y)$ for at least one (b, y) pair. We therefore have

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

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 $u(b, y, x, v_b)$. This is so if and only (i) (b, y) provides higher utility than that delivered by all within brand options $\{u(b, \tilde{y}, x, v_b) : \tilde{y} \neq y\}$, and (ii) (b, y) provides higher utility than that delivered by all alternative brand options $\left\{u\left(\tilde{b}, \tilde{y}, x, v_{\tilde{b}}\right) : \left(\tilde{b}, \tilde{y}\right) \neq (b, y)\right\}$.

Condition (i) requires that y maximizes $u(b, \cdot, x, v_b)$ 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 $g_b(\bar{y}_b) \equiv \infty$ 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 $u(b, y, x, v_b)$ 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

$$\begin{aligned} u(b, Y_b^*, x, v_b) &> \max_{y \in \mathcal{Y}_d} u(d, y, x, v_d), \text{ if } b < d, \\ u(b, Y_b^*, x, v_b) &\geq \max_{y \in \mathcal{Y}_d} u(d, y, x, v_d), \text{ if } b > d. \end{aligned}$$

By A6(ii) it follows that

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

is strictly monotonic and hence invertible in v_d. Therefore the above inequalities can be written as

$$\begin{aligned} g_d\{u(b, Y_b^*, x, v_b); x\} &> v_d, \text{ if } b < d, \\ g_d\{u(b, Y_b^*, x, v_b); x\} &\geq v_d, \text{ if } b > d, \end{aligned}$$

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

$$g_d(u_d^*(x, v_d); x) = v_d.$$

Then we have the inequalities (3.6) and (3.7) with

$$h_{bd}(y) \equiv g_d\{u(b, Y_b^*, x, v_b); x\},$$

for each pair $b \neq d$. The integral (3.4) for the conditional choice probabilities then follows immediately from their definition $p_{by}(x; \mathcal{S}) \equiv G(\mathcal{V}_{by}(x; u))$. ■

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

$$h(v, \theta) \equiv g(v) 1[v \in \mathcal{V}_{by}(x; \theta)]$$

is log-concave in (v, θ). This follows from log-concavity of $g(v)$ and log-concavity of $1[v \in \mathcal{V}_{by}(x; \theta)]$

in (v,θ), which is easy to establish given $\mathcal{V}_{by}(x;\theta)$ comprises a system of linear inequalities in (v,θ). By Theorem 6 of Prekopa (1973) it then follows that

$$\int_{\mathbb{R}^{|\mathcal{B}|}} h(v,\theta)\,dv$$

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 $b \in \{1,2\}$ let $\delta_b \equiv (\alpha_{b1},\beta_b')'$ and $\tilde{\delta}_b \equiv \left(\tilde{\alpha}_{b1},\tilde{\beta}_b'\right)'$. Identification of δ_1 and δ_2 follows directly from Aradillas-Lopez and Rosen (2014) Theorem 2. We provide the steps for completeness. Define the sets

$$S_b^+ \equiv \left\{z : z_1\left(\tilde{\delta}_1-\delta_1\right) > 0 \wedge z_2\left(\tilde{\delta}_2-\delta_2\right) \geq 0\right\},$$

$$S_b^- \equiv \left\{z : z_1\left(\tilde{\delta}_1-\delta_1\right) < 0 \wedge z_2\left(\tilde{\delta}_2-\delta_2\right) \leq 0\right\}.$$

For any $z \in S_b^+$ we have that

$$\Phi_2\left(z_1\tilde{\delta}_1, z_2\tilde{\delta}_2;\Sigma\right) > \Phi_2\left(z_1\delta_1, z_2\delta_2;\Sigma\right) = f_x^0(0), \tag{A.2}$$

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

$$\Phi_2\left(z_1\tilde{\delta}_1, z_2\tilde{\delta}_2;\Sigma\right) < \Phi_2\left(z_1\delta_1, z_2\delta_2;\Sigma\right) = f_x^0(0), \tag{A.3}$$

where Φ_2 denotes the cumulative distribution of a mean zero bivariate normal random variable with variance Σ, and where $f_x^0(0) = \mathbb{P}\{Y=0|X=x\}$. The probability that $Z \in S_b \equiv S_b^+ \cup S_b^-$ is

$$\begin{aligned}\mathbb{P}\{Z\in S_b\} &= \mathbb{P}\left\{Z\in S_b^+\right\} + \mathbb{P}\left\{Z\in S_b^-\right\}\\ &= \left(\begin{array}{c}\mathbb{P}\left\{Z_2\left(\tilde{\delta}_2-\delta_2\right)\geq 0|Z_1\left(\tilde{\delta}_1-\delta_1\right)>0\right\}\mathbb{P}\left\{Z_1\left(\tilde{\delta}_1-\delta_1\right)>0\right\}\\ +\mathbb{P}\left\{Z_2\left(\tilde{\delta}_2-\delta_2\right)\leq 0|Z_1\left(\tilde{\delta}_1-\delta_1\right)<0\right\}\mathbb{P}\left\{Z_1\left(\tilde{\delta}_1-\delta_1\right)<0\right\}\end{array}\right).\end{aligned}$$

Both $\mathbb{P}\left\{Z_1\left(\tilde{\delta}_1-\delta_1\right)>0\right\}$ and $\mathbb{P}\left\{Z_1\left(\tilde{\delta}_1-\delta_1\right)<0\right\}$ are strictly positive by condition (i), and at least one of $\mathbb{P}\left\{Z_2\left(\tilde{\delta}_2-\delta_2\right)\geq 0|Z_1\left(\tilde{\delta}_1-\delta_1\right)>0\right\}$ and $\mathbb{P}\left\{Z_2\left(\tilde{\delta}_2-\delta_2\right)\leq 0|Z_1\left(\tilde{\delta}_1-\delta_1\right)<0\right\}$ must be strictly positive by condition (ii). Therefore $\mathbb{P}\{Z\in S_b\}>0$, implying that $\tilde{\delta}$ is observationally distinct from δ since for each $z \in S_b$, $f_x^0(0) \neq \Phi_2\left(z_1\tilde{\delta}_1, z_2\tilde{\delta}_2;\Sigma\right)$. Thus

$\delta_1 = (\alpha_{11}, \beta_1')'$ and $\delta_2 = (\alpha_{21}, \beta_2')'$ are identified.

Given identification of (δ_1, δ_2), it is now straightforward to show that $(\alpha_{12}, \alpha_{22})$ are identified. Identification of (δ_1, δ_2) implies that in order for $\tilde{\theta}$ and θ to be observationally equivalent, we must have $\left(\tilde{\alpha}_{11}, \tilde{\alpha}_{21}, \tilde{\beta}_1, \tilde{\beta}_2\right) = (\alpha_{11}, \alpha_{21}, \beta_1, \beta_2)$. Suppose however that $\alpha_{12} > \tilde{\alpha}_{12}$. Then, if $p_{12}(x; \theta, \Sigma) = p_{12}\left(x; \tilde{\theta}, \Sigma\right)$ we must have that $\alpha_{22} > \tilde{\alpha}_{22}$, as otherwise $p_{12}(x; \theta, \Sigma) < p_{12}\left(x; \tilde{\theta}, \Sigma\right)$. But $\alpha_{12} > \tilde{\alpha}_{12}$ and $\alpha_{22} > \tilde{\alpha}_{22}$ together then imply that $p_{21}(x; \theta, \Sigma) > p_{21}\left(x; \tilde{\theta}, \Sigma\right)$ since the utility of choice $(b, y) = (2, 1)$ is the same for both $\tilde{\theta}$ and θ, but the utility of $(b, y) = (1, 2)$ or $(b, y) = (2, 2)$ is smaller at θ than at $\tilde{\theta}$. ■

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

Lemma 1 *When $\bar{y}_b = 2$ for each b, then (5.3) can be simplified to*

$$h_{by}(x, v, \theta) = 1\left[v_b < z_{by}^*\right] m_{by}^-(x, z, \theta) + 1\left[v_b \geq z_{by}^*\right] m_{by}^+(x, z, \theta), \tag{A.4}$$

where

$$m_{by}^-(x, v_b, \theta) \equiv y(x_b\beta_b + v_b) + \alpha_{d1} - \alpha_{by} - x_d\beta_d, \tag{A.5}$$

$$m_{by}^+(x, v_b, \theta) \equiv \frac{1}{2}\left[y(x_b\beta_b + v_b) + \alpha_{d2} - \alpha_{by}\right] - x_d\beta_d, \tag{A.6}$$

Proof. Since $\bar{y}_b = 2$, (5.3) simplifies to

$$h_{by}(x, v, \theta) = \min_{\tilde{y} \in \{1, \ldots, \bar{y}_d\}} \frac{1}{\tilde{y}}\left[y(x_b\beta_b + v_b) - (\alpha_{by} - \alpha_{d\tilde{y}})\right] - x_d\beta_d \tag{A.7}$$

$$= \min\left\{m_{by}^-(x, v_b, \theta), m_{by}^+(x, v_b, \theta)\right\}. \tag{A.8}$$

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

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

and since $m_{by}^-(x, v_b, \theta)$ has a larger slope with respect to v_b, it follows that for all $v_b < z_{by}^*$, $m_{by}^-(x, v_b, \theta) < m_{by}^+(x, v_b, \theta)$, while for all $v_b > z_{by}^*$, $m_{by}^-(x, v_b, \theta) > m_{by}^+(x, v_b, \theta)$. Thus (A.8) simplifies to (A.4), completing the proof. ■

Proof of Proposition 2. The starting point is (5.4):

$$p_{by}(x,\theta) = \frac{1}{\sigma_b} \int_{\lambda_{b,y}}^{\lambda_{b,y+1}} \Phi\left(\frac{h_{by}(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi\left(\frac{z}{\sigma_b}\right) dz,$$

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

1. $\lambda_{b,y} < z^*_{by} < \lambda_{b,y+1}$.

$$p_{by}(x,\theta) = \sigma_b^{-1} \int_{\lambda_{b,y}}^{z^*_{by}} \Phi\left(\frac{m^-_{by}(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi\left(\frac{z}{\sigma_b}\right) dz + \sigma_b^{-1} \int_{z^*_{by}}^{\lambda_{b,y+1}} \Phi\left(\frac{m^+_{by}(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi\left(\frac{z}{\sigma_b}\right) \tag{A.9}$$

2. $\lambda_{b,y} \leq \lambda_{b,y+1} \leq z^*_{by}$.

$$p_{by}(x,\theta) = \sigma_b^{-1} \int_{\lambda_{b,y}}^{\lambda_{b,y+1}} \Phi\left(\frac{m^-_{by}(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi\left(\frac{z}{\sigma_b}\right) dz \tag{A.10}$$

3. $z^*_{by} \leq \lambda_{b,y} \leq \lambda_{b,y+1}$.

$$p_{by}(x,\theta) = \sigma_b^{-1} \int_{\lambda_{b,y}}^{\lambda_{b,y+1}} \Phi\left(\frac{m^+_{by}(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi\left(\frac{z}{\sigma_b}\right) dz \tag{A.11}$$

The expressions in each case simplify as follows, using (A.5) and (A.6) and a change of variables substitution for $\frac{z}{\sigma_b}$.

$$\sigma_b^{-1} \int_{\lambda_{b,y}}^{z^*_{by}} \Phi\left(\frac{m^-_{by}(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi\left(\frac{z}{\sigma_b}\right) dz$$

$$= \int_{\sigma_b\lambda_{b,y}}^{\sigma_b z^*_{by}} \Phi\left(\frac{yx_b\beta_b + \alpha_{d1} - \alpha_{by} - x_d\beta_d + (\sigma_b y - \rho\sigma_d)z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi(z)\, dz \tag{A.12}$$

$$\sigma_b^{-1} \int_{z_{by}^*}^{\lambda_{b,y+1}} \Phi\left(\frac{m_{by}^+(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{z}{\sigma_b}\right)$$

$$= \int_{\sigma_b z_{by}^*}^{\sigma_b\lambda_{b,y+1}} \Phi\left(\frac{\frac{1}{2}\left[yx_b\beta_b + \alpha_{d2} - \alpha_{by}\right] - x_d\beta_d + \left(\frac{1}{2}\sigma_b y - \rho\sigma_d\right)z}{\sigma_d\sqrt{1-\rho^2}}\right)\phi(z)\,dz. \quad \text{(A.13)}$$

$$\sigma_b^{-1} \int_{\lambda_{b,y}}^{\lambda_{b,y+1}} \Phi\left(\frac{m_{by}^-(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{z}{\sigma_b}\right)dz$$

$$= \int_{\sigma_b\lambda_{b,y}}^{\sigma_b\lambda_{b,y+1}} \Phi\left(\frac{yx_b\beta_b + \alpha_{d1} - \alpha_{by} - x_d\beta_d + (\sigma_b y - \rho\sigma_d)z}{\sigma_d\sqrt{1-\rho^2}}\right)\phi(z)\,dz \quad \text{(A.14)}$$

$$\sigma_b^{-1} \int_{\lambda_{b,y}}^{\lambda_{b,y+1}} \Phi\left(\frac{m_{by}^+(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{z}{\sigma_b}\right)dz$$

$$= \int_{\sigma_b\lambda_{b,y}}^{\sigma_b\lambda_{b,y+1}} \Phi\left(\frac{\frac{1}{2}\left[yx\beta_b + \alpha_{d2} - \alpha_{by}\right] - x_d\beta_d + \left(\frac{1}{2}\sigma_b y - \rho\sigma_d\right)z}{\sigma_d\sqrt{1-\rho^2}}\right)\phi(z)\,dz \quad \text{(A.15)}$$

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

$$\int_h^k \Phi(c_1 + c_2 z)\,\phi(z)\,dz = \Lambda(k, c_1, c_2) - \Lambda(h, c_1, c_2) \quad \text{(A.16)}$$

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

$$\Lambda(k, c_1, c_2) = \int_{-\infty}^{\frac{c_1}{\sqrt{c_2^2+1}}} \phi(z)\,\Phi\left(k\sqrt{c_2^2+1} + c_2 z\right)dz. \quad \text{(A.17)}$$

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

$$\int_{-\infty}^{y} \phi(z)\,\Phi(a+bz)\,dz = \Phi_2\left(\frac{a}{\sqrt{1+b^2}}, y; \frac{-b}{\sqrt{1+b^2}}\right). \tag{A.18}$$

Applying this formula to (A.17) with

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

gives

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

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

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

Referring back to (A.16), substitution of c_2 with those coefficients multiplying z and substitution of c_1 with those terms not multiplying z in the integrands on the rights hand side of (A.12) - (A.15) combined with (A.9) - (A.11) gives the following expression for conditional choice probabilities according to where z_{by}^* lies with respect to the interval $[\lambda_{b,y}, \lambda_{b,y+1}]$.

1. $\lambda_{b,y} < z_{by}^* < \lambda_{b,y+1}$.

$$\begin{aligned} p_{by}(x,\theta) &= \sigma_b^{-1}\left\{ \begin{array}{c} \int_{\lambda_{b,y}}^{z_{by}^*} \Phi\left(\frac{m_{by}^-(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{z}{\sigma_b}\right)dz \\ + \int_{z_{by}^*}^{\lambda_{b,y+1}} \Phi\left(\frac{m_{by}^+(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b}z}{\sigma_d\sqrt{1-\rho^2}}\right)\phi\left(\frac{z}{\sigma_b}\right) \end{array} \right\} \\ &= \tilde{\Delta}\left(\sigma_b^{-1}\lambda_{b,y}, \sigma_b^{-1}z_{by}^*, c_1^-, c_2^-\right) + \tilde{\Delta}\left(\sigma_b^{-1}z_{by}^*, \sigma_b^{-1}\lambda_{b,y+1}, c_1^+, c_2^+\right). \end{aligned}$$

2. $\lambda_{b,y} \leq \lambda_{b,y+1} \leq z^*_{by}$.

$$\begin{aligned} p_{by}(x,\theta) &= \sigma_b^{-1} \int_{\lambda_{b,y}}^{\lambda_{b,y+1}} \Phi\left(\frac{m^-_{by}(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b} z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi\left(\frac{z}{\sigma_b}\right) dz \\ &= \tilde{\Delta}\left(\sigma_b^{-1}\lambda_{b,y}, \sigma_b^{-1}\lambda_{b,y+1}, c_1^-, c_2^-\right). \end{aligned}$$

3. $z^*_{by} \leq \lambda_{b,y} \leq \lambda_{b,y+1}$.

$$\begin{aligned} p_{by}(x,\theta) &= \sigma_b^{-1} \int_{\lambda_{b,y}}^{\lambda_{b,y+1}} \Phi\left(\frac{m^+_{by}(x,z,\theta) - \rho\frac{\sigma_d}{\sigma_b} z}{\sigma_d\sqrt{1-\rho^2}}\right) \phi\left(\frac{z}{\sigma_b}\right) dz \\ &= \tilde{\Delta}\left(\sigma_b^{-1}\lambda_{b,y}, \sigma_b^{-1}\lambda_{b,y+1}, c_1^+, c_2^+\right). \end{aligned}$$

Using indicators for whether $z^*_{by} < \lambda_{b,y+1}$ and $z^*_{by} > \lambda_{b,y}$ to cover each of these cases gives

$$p_{by}(x,\theta) = \begin{pmatrix} 1\left[z^*_{by} < \lambda_{b,y+1}\right] \tilde{\Delta}\left(\sigma_b^{-1}\max\left\{z^*_{by}, \lambda_{b,y}\right\}, \sigma_b^{-1}\lambda_{b,y+1}, c_1^+, c_2^+\right) \\ +1\left[z^*_{by} > \lambda_{b,y}\right] \tilde{\Delta}\left(\sigma_b^{-1}\lambda_{b,y}, \sigma_b^{-1}\min\left\{z^*_{by}, \lambda_{b,y+1}\right\}, c_1^-, c_2^-\right) \end{pmatrix}. \tag{A.19}$$

This produces (5.5) by noting that that variables defined in (5.7) and (5.8) 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 that $p_{by}(x,\theta)$ in (A.19) is equal to (5.5) in the statement of the Theorem 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 the UK. The razor blade market is divided into three different sectors: cartridges bought with a razor, cartridges bought alone, known as "system blades", and disposable razors. The original data consists of 7234 observations. Table 9 shows the market share of the three sectors in 2004-2005.

Sector	Total
Bought with razor	16.80%
System blades	33.67%
Disposable razors	49.53%

Table 9: Market shares of the sectors Bought with razor, 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 an already existing handle. We define the outside option as buying a disposable blade. Disposable blades are on average sold at a cheaper price relative to reusable cartridges, which can indicate that conditional on the blade type, disposable razors are considered of lower quality. On average in our sample and using equation (6.1), double blade disposable razors cost £0.26 and a triple blade razors cost £0.63. For system blades a double blade cartridge costs £0.79 and a triple blade cartridge costs £1.45, on average.[13]

The market for reusable razors is dominated by two firms, Gillette and Wilkinson Sword, each offering razors and cartridges with two or three blades.[14] Gillette's double blade reusable razor model, Sensor, was introduced in 1992 and the triple blade reusable razor model, Venus, was introduced in 2001. Wilkinson Sword introduced its double blade reusable razor, Lady Protector, in 1994, while its triple blade reusable model, Intuition, was introduced in 2003.

We use observations 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.[15] In the analysis we also include the marital status of the main shopper, and a variable indicating whether there is more than one female in the household. Table 10 gives summary statistics of the main shopper characteristics.

For each individual 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

[13]According to Gillette, Venus disposable razors are for one-time or limited use and in general last between three to ten shaves, while reusuable razor cartridges typically last five to ten shaves, and are also environmentally friendlier (source: www.gillettevenus.com/en-us/womens-shaving-guide/learning-to-shave/disposable-razors-vs-refillable-razors/)

[14]Some stores offer own-label reusable razors but these were dropped from the sample as they accounted for only 4.41% of the market share for system blades.

[15]Retired individuals were excluded from the sample.

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	38.17%	More than one	56.96%
Works less than 8 hours	2.38%				
Unemployed/not working	28.13%				
Full time education	1.12%				

Table 10: Main shopper characteristics for 2004-2005.

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 Tables 11 and 12 , cartridges of system blades were offered 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 cartidges, medium (M) if they contained 5 or 6 cartridges, and large (L) if they contained 8 or more cartridges.

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 11: Pack sizes offered.

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

Table 12: Pack size grouping.

The counterfactual prices in equations (6.2) and (6.3) were calculated both conditioning and not conditioning on the pack size of the product purchased, respectively. As is evident from Table 11 not all blade types and not all brands offer all pack sizes. Table 14 gives the estimates of regressions (6.2) and (6.3). For the calculation of the average price in equation (6.1) and of the counterfactual prices in equations (6.2) and (6.3), the individual shops were grouped using the company groups in Table 13.

Company group	Total
Asda	22.24%
Boots	8.57%
Co-op	0.48%
Kwiksave	0.56%
Morrisons	7.95%
Safeway	1.47%
Sainsbury	7.97%
Savacentre	0.64%
Somerfield	0.91%
Superdrug	3.90%
Tesco	25.20%
Waitrose	0.33%
Wilkinson	13.78%
Default	3.61%
All other	2.40%

Table 13: Company groups of stores observed.

	Specification (6.2)	Specification (6.3)
Wilkinson Sword	0.4048***	0.3961***
	(0.008)	(0.0081)
Triple blade	0.5043***	0.6948***
	(0.024)	(0.0082)
Medium pack size	-0.1988***	
	(0.0236)	
Large pack size	-0.0205	
	(0.0277)	
Constant	0.7385***	0.5456***
	(0.0429)	(0.037)

Notes: Monthly and store dummies are suppressed. Standard errors are reported in parentheses. *** denotes significance at 1%.

Table 14: OLS Regression Estimates of regressions (6.2) and (6.3)