The Empirical Content of Binary Choice Models*

Debopam Bhattacharya†

University of Cambridge

November 12, 2018

Abstract

Empirical demand models used for counterfactual predictions and welfare analysis must be rationalizable, i.e. theoretically consistent with utility maximization by heterogeneous consumers. We show that for binary choice under general unobserved heterogeneity, rationalizability is equivalent to a pair of Slutsky-like shape-restrictions on choice-probability functions. The forms of these restrictions differ from Slutsky-inequalities for continuous goods. Unlike McFadden-Richter's stochastic revealed preference, our shape-restrictions (a) are global, i.e. their forms do not depend on which and how many budget-sets are observed, (b) are closed-form, hence easy to impose on parametric/semi/non-parametric models in practical applications, and (c) provide computationally simple, theory-consistent bounds on demand and welfare predictions on counterfactual budget-sets.

1 Introduction

Many important economic decisions faced by individuals are binary in nature, including labour force participation, retirement, college enrolment, adoption of a new technology or health product, and so forth. This paper concerns nonparametric analysis of binary choice under general unobserved heterogeneity and income effects. The paper has two goals. The first is to understand, theoretically, what nonparametric restrictions does utility maximization by heterogeneous consumers impose upon choice-probabilities, i.e. whether there are analogs of Slutsky restrictions for binary

*Keywords: Binary choice, general heterogeneity, income effect, utility maximization, integrability/rationalizability, Slutsky inequality, shape-restrictions. JEL Codes: C14, C25, D12.

†The author would like to thank Arthur Lewbel, Oliver Linton and seminar participants at several institutions for helpful feedback. Financial support from the European Research Council via a Consolidator Grant EDWEL is gratefully acknowledged.

choice under general unobserved heterogeneity and income effects, and conversely, whether these restrictions are also sufficient for observed choice-probabilities to be rationalizable. This issue is important for logical coherency between theory and empirics and, in particular, for prediction of demand and welfare in situations involving counterfactual, i.e. previously unobserved, budget sets. It is important in these exercises to allow for general unobserved heterogeneity, because economic theory typically does not restrict its dimension or distribution, and does not specify how it enters utility functions. To date, closed-form Slutsky conditions for rationalizability of demand under general heterogeneity were available only for continuous choice. The present paper, to our knowledge, is the first to establish them for the leading case of discrete demand, viz. binary choice.

The second goal of the present paper is a practical one. It is motivated by the fact that in empirical applications of binary choice, requiring the estimation of elasticities, welfare calculations and demand predictions, researchers typically use parsimonious functional-forms for conditional choice probabilities. This is because fully nonparametric estimation is often hindered by curse of dimensionality, the sensitivity of estimates to the choice of tuning parameters and insufficient price variation, especially in consumer data from developed countries. The question therefore arises as to whether the economic theory of consumer behavior can inform the choice of such functional forms. Answering this question is our second objective.

Since McFadden, 1973, discrete choice models of economic behavior have been studied extensively in the econometric literature, mostly under restrictive assumptions on utility functions and unobserved heterogeneity including, inter alia, quasi-linear preferences implying absence of income effects and/or parametrically specified heterogeneity distributions (c.f. Train, 2009 for a textbook treatment). Matzkin (1992) investigated the nonparametric identification of binary choice models with additive heterogeneity, where both the distribution of unobserved heterogeneity and the functional form of utilities were left unspecified. More recently, Bhattacharya (2015, 2018) has shown that in discrete choice settings, welfare distributions resulting from price changes are nonparametrically point-identified from choice probabilities without any substantive restriction on preference heterogeneity, and even when preference distribution and heterogeneity dimension are not identified.

In the present paper, we consider a setting of binary choice by a population of budget-constrained consumers with general, unobserved heterogeneity. In this setting, we develop a characterization of utility maximization which takes the form of simple, closed-form shape restrictions on choice probability functions in the population. These nonparametric shape-restrictions can be consistently tested in the usual asymptotic econometric sense and are extremely easy to impose on specifications of choice-probabilities – akin to testing or imposing monotonicity of regression functions. Most importantly, they lead to computationally simple bounds for theory-consistent demand and welfare predictions on counterfactual budgets sets – an important goal of empirical demand analysis. Interestingly, our shape-restrictions differ in form from the well-known Slutsky inequalities for continuous goods.

The above results are developed in a fully nonparametric context; nonetheless, they can help guide applied researchers intending to use simple parametric or semiparametric models. As a specific example, consider the popular probit/logit type model for binary choice of whether to buy a product or not. A standard specification is that the probability of buying depends (implicitly conditioning on other observed covariates) on its price p and the decision-maker's income y, e.g. $\bar{q}(p, y) = F(\gamma_0 + \gamma_1 p + \gamma_2 y)$, where $F(\cdot)$ is a distribution function. We will show below that these choice-probabilities are consistent with utility maximization by a heterogenous population of consumers, if and only if $\gamma_1 \leq 0$, and $\gamma_1 + \gamma_2 \leq 0$. While the first inequality simply means that demand falls with own price (holding income fixed), the second inequality is less obvious, and constitutes an important empirical characterization of utility maximization.

For the case of *continuous* goods, Lewbel, 2001 explored the question of when average demand, generated from maximization of heterogeneous individual preferences, satisfies standard properties of non-stochastic demand functions. More recently, for the case of two continuous goods (i.e. a good of interest plus the numeraire) under general heterogeneity, Hausman and Newey, 2016 have shown that constrained utility maximization is equivalent to quantiles of demand satisfying standard Slutsky negativity. The analog of the two goods setting in discrete choice is the case of binary alternatives. Accordingly, our main result (Theorem 1 below) may be viewed as the discrete choice counterpart of Hausman and Newey, 2016, Theorem 1. Note however that quantiles are degenerate for binary outcomes, and indeed, the forms of our Slutsky-like shape restrictions are completely different from Hausman-Newey's quantile-based conditions for continuous choice.

An alternative, algorithmic – as opposed to closed-form and analytic – approach to rationalizability of demand is the so-called "revealed stochastic preference" (SRP, henceforth) method, which applies to very general choice settings where a heterogeneous population of consumers faces a finite number of budget sets, c.f. McFadden and Richter, 1990, McFadden, 2005. When budget sets are numerous or continuously distributed, as in household surveys with many income and/price values, SRP is well-known to be operationally prohibitive, c.f. Anderson et al 1992, Page 54-5 and Kitamura and Stoye, 2016, Sec 3.3. Furthermore, the SRP conditions are difficult to impose on parametric specifications commonly used in practical applications, they change entirely in form upon addition of new budget sets, and are cumbersome to use for demand prediction on counterfactual budgets, especially in welfare calculations that typically require simultaneous prediction on a continuous range of budget-sets. In contrast, our approach yields rationality conditions which (a) are global, in that they characterize choice probability *functions*, and their forms remain invariant to which and how many budget sets are observed in a dataset, and (b) are closed-form, analytic shape-restrictions, hence easy to impose, standard to test, and simple to use for counterfactual predictions of demand and welfare. As such, these shape-restrictions establish the analogs of Slutsky conditions – the cornerstone of classical demand analysis – for binary choice under general unobserved heterogeneity and income effects.

2 The Result

Consider a population of heterogeneous individuals, each choosing whether or not to buy an indivisible good. Let N represent the quantity of numeraire which an individual consumes in addition to the binary good. If the individual has income $Y = y$, and faces a price $P = p$ for the indivisible good, then the budget constraint is $N + pQ = y$ where $Q \in \{0, 1\}$ represents the binary choice. Individuals derive satisfaction from both the indivisible good as well as the numeraire. Upon buying, an individual derives utility from the good but has a lower amount of numeraire $y - p$ left; upon not buying, she enjoys utility from her outside option and a higher quantity of numeraire y. There is unobserved heterogeneity across consumers which affect their choice, and so on each budget set defined by a price p and consumer income y, there is a (structural) probability of buying, denoted by $q(y, y-p)$; that is, if each member of the entire population were offered income y and price p, then a fraction $q(y, y-p)$ would buy the good. It is more standard to write this choice probability as conditional on price and income, i.e., in the form $\bar{q}(p, y)$, but the equivalent $q(y, y-p)$ expression is an important and helpful step toward obtaining closed-form rationalizability results, as will become clear below. Indeed, one can go back and forth between the two specifications because $\bar{q}(c, d) \equiv q(d, d-c)$ and $q(a, b) \equiv \bar{q}(a-b, a)$. Also, for now, we implicitly condition our analysis on observed covariates, and later show how to incorporate them into the results.

Our main result establishes conditions that are necessary and sufficient for the conditional choice probability function to be generated from utility maximization by a heterogeneous population, where no a priori restriction is imposed on the dimension and functional form of the distribution of unobserved heterogeneity or on the functional form of utilities.

Theorem 1 *For binary choice under general heterogeneity, the following two statements are equivalent:*

(i) The structural choice probability function $q(\cdot,\cdot)$, defined above, satisfies that (A) $q(a,b)$ is non-increasing in a for each fixed b, and non-decreasing in b for each fixed a; (B) for each fixed $b \in \mathbb{R}$, it holds that $\lim_{a\downarrow-\infty} q(a,b) = 1$, and (C) $q(a,b)$ is continuous in a for each fixed b.

(ii) There exists a pair of utility functions $W_0(\cdot,\eta)$ and $W_1(\cdot,\eta)$, where the first argument denotes the amount of numeraire, and η denotes unobserved heterogeneity, and a distribution $G(\cdot)$ of η such that

$$q(a,b) = \int 1\{W_0(a,\eta) \leq W_1(b,\eta)\}\, dG(\eta),$$

where (A') for each fixed η, $W_0(a,\eta)$ is continuous and strictly increasing in a, and $W_1(b,\eta)$ is non-decreasing in b; (B') for each fixed b and η, $W_1(b,\eta) > \lim_{a\downarrow-\infty} W_0(a,\eta)$; (C') for any a,b, it holds that $\int 1\{W_1(b,\eta) = W_0(a,\eta)\}\, dG(\eta) = 0$.

Intuitively, conditions (A/A') mean that having more numeraire ceteris paribus is (weakly) better for every consumer. To interpret condition (B), note that for fixed b, $\lim_{a\downarrow-\infty} q(a,b) \equiv \lim_{(a-b)\downarrow-\infty} q(a,b)$. Now, since $a-b=p$ is the price, condition (B/B') say that everyone can be persuaded to buy product 1 by making its price low enough (perhaps even negative). Condition (C/C') – the "no-tie" assumption – is standard in discrete choice models, and intuitively means that there is a continuum of tastes. Note that (A)-(C) place *no restriction* on income effects, including its sign.

In statement (ii), the functions $W_j(x,\eta)$ will correspond to the utility from choosing alternative $j \in \{0,1\}$ and being left with a quantity x of the numeraire, and with η denoting unobserved heterogeneity. This notation allows for the case where different vectors of unobservables enter the two utilities, i.e. where the utilities are given by $u_0(\cdot,\eta_0)$ and $u_1(\cdot,\eta_1)$, respectively, with $\eta_0 \neq \eta_1$; simply set $\eta \equiv (\eta_0,\eta_1)$, $W_0(\cdot,\eta) \equiv u_0(\cdot,\eta_0)$, $W_1(\cdot,\eta) \equiv u_1(\cdot,\eta_1)$. In the proof of the above theorem, when showing (ii) implies (i), η will be allowed to have *any arbitrary and unknown* dimension and distribution; in showing (i) implies (ii) we will construct a scalar heterogeneity distribution that will rationalize the choice probabilities (see further discussion on this point under the heading "Observational Equivalence" in the next section).

Proof. That (ii) implies (i) is straightforward. In particular, letting $W_0^{-1}(\cdot,\eta)$ denote the inverse of $W_0(\cdot,\eta)$, we have that

$$q(a,b)=\int 1\left\{a\leq W_0^{-1}\left(W_1(b,\eta),\eta\right)\right\}dG(\eta)$$

whence (C') implies continuity of $q(\cdot,b)$, (B') implies that $\lim_{a\downarrow-\infty}q(a,b)=1$ for each b, and (A') implies (A).

We now show that (i) implies (ii). Note that $\lim_{a\downarrow-\infty}q(a,b)=1$ for each b implies that for any $u\in[0,1]$ and $b\in\mathbb{R}$, the set $\{x:q(x,b)\geq u\}$ is non-empty. For any fixed $b\in\mathbb{R}$ and for $u\in[0,1]$, define

$$q^{-1}(u,b)\overset{def}{=}\sup\{x:q(x,b)\geq u\}. \tag{1}$$

By condition (A) in (i), $q^{-1}(u,\cdot)$ must be non-decreasing. Now, consider a random variable $V\simeq Uniform(0,1)$. Define $W_0(a,V)\overset{defn}{=}a$ and $W_1(b,V)\overset{defn}{=}q^{-1}(V,b)$, and note that by construction, $W_0(a,V)$ and $W_1(b,V)$ satisfy properties (A')-(C') listed in (ii) above. We will now show that $W_0(\cdot,V)$ and $W_1(\cdot,V)$ will rationalize the choice-probabilities $q(\cdot,\cdot)$.

Indeed, given any fixed b, since $q(\cdot,b)$ is continuous and non-increasing, we have that for any $v\in(0,1)$,

$$a\leq q^{-1}(v,b)\overset{\text{by }q(\cdot,b)\ non\uparrow}{\Longrightarrow}q(a,b)\geq q\left(q^{-1}(v,b),b\right)\overset{\text{by }q(\cdot,b)\ cont.}{\Longrightarrow}q(a,b)\geq v. \tag{2}$$

To see why continuity is required for the last implication in (2), suppose for some $v\in(0,1)$, we have that $q(x,b)>v$ for all $x<c$, but $q(c,b)<v$, i.e. $q(\cdot,b)$ takes a discontinuous 'plunge' at c. Then $q^{-1}(v,b)=\sup\{x:q(x,b)\geq v\}=c$, but $q(c,b)=q\left(q^{-1}(v,b),b\right)<v$. Continuity of $q(\cdot,b)$ rules this out, and guarantees that $q(c,b)=q\left(q^{-1}(v,b),b\right)\geq v$; therefore, in (2), $q(a,b)\geq q\left(q^{-1}(v,b),b\right)\Longrightarrow q(a,b)\geq v$.

Finally, by definition of $q^{-1}(\cdot,b)$ as the supremum in (1), we have that

$$q(a,b)\geq v\Longrightarrow a\leq q^{-1}(v,b). \tag{3}$$

Therefore, by (2) and (3), we have that $q(a,b)\geq v\Longleftrightarrow a\leq q^{-1}(v,b)$, and thus, for $V\simeq U(0,1)$, it follows that

$$\Pr\left(q^{-1}(V,b)\geq a\right)=\Pr\left(V\leq q(a,b)\right)=q(a,b).$$

Therefore, the utility functions $W_0(a,V)\equiv a$ and $W_1(b,V)\equiv q^{-1}(V,b)$ with heterogeneity $V\simeq Uniform(0,1)$ rationalize the choice probabilities $q(\cdot,\cdot)$, and satisfy all the properties specified in panel (ii) of Theorem 1. In particular, $W_1(b,\eta)$ is non-decreasing in b (see (1)). ■

3 Discussion

A. Slutsky Form: To see the analogy between the shape restrictions in Theorem 1 and the traditional Slutsky inequality constraints with smooth demand, rewrite the choice probability on a budget set (p, y) in the standard form as a function of price and income, viz. $\bar{q}(p, y) \equiv q(y, y - p)$ i.e., $q(a, b) \equiv \bar{q}(a - b, a)$. Then, under continuous differentiability, the shape restrictions (A) from Theorem 1 are equivalent to $\frac{\partial}{\partial b}\bar{q}(a - b, a) = \frac{\partial}{\partial b}q(a, b) \geq 0$, and $\frac{\partial}{\partial a}\bar{q}(a - b, a) = \frac{\partial}{\partial a}q(a, b) \leq 0$, i.e., for all p, y,

$$\frac{\partial}{\partial p}\bar{q}(p, y) \leq 0, \tag{4}$$

$$\frac{\partial}{\partial p}\bar{q}(p, y) + \frac{\partial}{\partial y}\bar{q}(p, y) \leq 0. \tag{5}$$

The forms of these inequalities are distinct from textbook Slutsky conditions for *nonstochastic* demand $q^*(p, y)$ for a *continuous* good, which are given by

$$\frac{\partial}{\partial p}q^*(p, y) + q^*(p, y)\frac{\partial}{\partial y}q^*(p, y) \leq 0 \text{ for all } p, y. \tag{6}$$

For a continuous good and under general unobserved heterogeneity, Hausman and Newey, 2016 show that (6) also holds with $q^*(p, y)$ denoting any quantile of the demand distribution for fixed (p, y) (see also Dette, Hoderlein and Neumeyer, 2016). Thus, for binary choice with general heterogeneity, the forms of the Slutsky inequality (4) and (5) are different from the continuous choice counterpart (6).[1] In particular, the inequalities (4) and (5) are *linear* in $\bar{q}(\cdot, \cdot)$ (and $q(\cdot, \cdot)$), unlike (6), and hence easier to impose on nonparametric estimates of $q(\cdot, \cdot)$ using, say, shape-preserving sieves that guarantee that $\frac{\partial}{\partial b}\hat{q}(a, b) \geq 0$, and $\frac{\partial}{\partial a}\hat{q}(a, b) \leq 0$ for all a, b.

Remark 1 *It is tempting to think of (4) and (5) as (6) with the level $q^*(p, y)$ replaced by 0 and 1 corresponding to either of the two possible individual choices. However, this interpretation is incorrect, since $\bar{q}(p, y)$ is* ***average*** *demand, and takes values strictly inside* $(0, 1)$*. In other words, $\bar{q}(p, y)$ is neither a quantile, nor individual demand at price p and y, and generically (e.g. in a probit model) does not take the values of 0 and 1. Thus (4) and (5)* ***cannot*** *be rewritten as*

$$\frac{\partial}{\partial p}\bar{q}(p, y) + \bar{q}(p, y)\frac{\partial}{\partial y}\bar{q}(p, y) \leq 0 \textit{ for all } p, y,$$

and, as such, are different from the continuous choice counterpart (6).

[1] Bhattacharya, 2015 (see also Lee and Bhattacharya, 2018) noted that (4) (resp, (5)) is necessary for the CDF of equivalent variation (resp., compensating variation) resulting from price-changes to be non-decreasing.

B. Observational Equivalence: The construction in our proof of (ii) $\Rightarrow$ (i) shows that a rationalizable binary choice model with general heterogeneity of unspecified dimension is observationally equivalent to one where a scalar heterogeneity enters the utility function of one of the alternatives in a monotonic way, and the utility of the other alternative is non-stochastic.[2] An intuitive explanation of this equivalence is that in the binary case, choice probabilities are determined solely by the marginal distribution of reservation price (given income) for alternative 1, and not the relative ranking of individual consumers in terms of their preferences within that distribution. So, as income varies, choice probabilities change only insofar as the marginal distribution of the reservation price changes, irrespective of how individual consumers' relative positions change within that distribution.

It is worth pointing out here that a binary choice model with *additive* scalar heterogeneity – the so-called ARUM model – is restrictive, and *not* observationally equivalent to a binary choice model with general heterogeneity. To see this, suppose choice probabilities are generated via the ARUM model, viz.

$$\begin{aligned} q(a,b) &= \Pr\left[W_1(b)+\eta_1 > W_0(a)+\eta_0\right] \\ &= \Pr\left[\eta_0-\eta_1 < W_1(b)-W_0(a)\right] \\ &= F_{\eta_0-\eta_1}\left[W_1(b)-W_0(a)\right]. \end{aligned} \tag{7}$$

Assuming smoothness and strict monotonicity of $F_{\eta_0-\eta_1}[\cdot]$, $W_1(\cdot)$ and $W_0(\cdot)$, and thus of $q(\cdot,\cdot)$, it follows that

$$\begin{aligned} & \frac{\partial^2}{\partial a\partial b}\ln\left[-\frac{\frac{\partial}{\partial b}q(a,b)}{\frac{\partial}{\partial a}q(a,b)}\right] \\ = & \frac{\partial^2}{\partial a\partial b}\ln\left(\frac{W_1'(b)}{W_0'(a)}\right), \text{ from (7)} \\ = & \frac{\partial^2}{\partial a\partial b}\left[\ln\left(W_1'(b)\right)-\ln\left(W_0'(a)\right)\right] \\ = & 0, \end{aligned}$$

for every a and b. This equality is obviously not true for a general smooth and strictly monotone $q(\cdot,\cdot)$ satisfying Assumptions i(A)-i(C) of Theorem 1.

[2] For quantile demand in the continuous case, a result of similar spirit is discussed in Hausman-Newey, 2016, Page 1228-9, following Theorem 1. In general, a result holding for the continuous case with two goods does not necessarily imply that it also holds for the binary case. For example, welfare related results are different for the binary and the two-good continuous case, c.f. Hausman-Newey 2016, and Bhattacharya 2015, and so are Slutsky negativity conditions, as discussed above.

Remark 2 *The construction of $q^{-1}(V,\cdot)$ in our proof of (ii) $\Rightarrow$ (i) is unrelated to the almost sure representation of a continuous random variable X as $F_X^{-1}(U)$ with $U = F_X(X)$, where F_X and F_X^{-1} denote the CDF and quantile function of X, and U is $U(0,1)$. Indeed, if we were to apply this so-called "probability-integral transform" to $X = W_1(a_1,\eta)$ for a fixed a_1, we will have $W_1(a_1,\eta) \overset{a.s.}{=} F_{W_1(a_1,\eta)}^{-1}(U(a_1))$, where the scalar-valued uniform process $U(a_1) \equiv F_{W_1(a_1,\eta)}(W_1(a_1,\eta))$ will vary with a_1, unlike V in the proof of our theorem above, and therefore cannot represent unobserved heterogeneity in consumer preferences. In other words, our constructed $q^{-1}(V,a_1)$ will not equal the data generating process $W_1(a_1,\eta)$ almost surely, but the probability that $q^{-1}(V,a_1) \geq a_0$ will equal the probability that $W_1(a_1,\eta) \geq W_0(a_0,\eta)$ for all (a_0,a_1).*

C. Giffen Goods: Our rationalizability condition (4) says that own price effect on average demand is negative. This condition has no counterpart in the continuous case, appears to rule out Giffen behavior and may, therefore, appear restrictive. We now show that that is not the case: indeed, Giffen goods cannot arise in binary choice models if utilities are non-satiated in the numeraire. To see this, let the utility of options 0 and 1 be given by $W_0(\cdot,\eta)$ and $W_1(\cdot,\eta)$ as in Theorem 1 above. Now note that if option 1 is Giffen for an η type consumer with income y, then for some prices $p < p'$ she buys at price p' but does not buy at p. Therefore,

$$W_1(y-p,\eta) < W_0(y,\eta) < W_1\left(y-p',\eta\right),$$

which is a contradiction, since $W_1(\cdot,\eta)$ is strictly increasing. In contrast, consider a *continuous* good with utilities $W(x, y-px, \eta)$, where x denotes the quantity of the continuous good, and $W(\cdot,\cdot,\eta)$ is increasing in both arguments. Now it is possible that x is bought at price p and x' is bought at price p' with $p < p'$ and $x < x'$. That is, we can have

$$W(x, y-px, \eta) < W\left(x', y-p'x', \eta\right),$$

if x' is preferred sufficiently over x. The intuitive reason for this difference between the discrete and the continuous case is that in the former, the only non-zero option is 1. Indeed, in the continuous case, it is also not possible that $W(x, y-px, \eta) < W(x, y-p'x, \eta)$ for any *common* x if $p < p'$.

Also, note that although Giffen behavior cannot arise in binary choice, there is no restriction on the *sign of the income effect*. Indeed, (4) and (5) are compatible with both $\frac{\partial}{\partial y}\bar{q}(p,y) \geq 0$ and $\frac{\partial}{\partial y}\bar{q}(p,y) \leq 0$.

D. Parametric and Semiparametric Models: For a probit/logit specification of the buying decision, viz.

$$\begin{aligned} & \bar{q}(p, y) \\ = \; & F(\gamma_0 + \gamma_1 p + \gamma_2 y) \\ = \; & F(\gamma_0 + (\gamma_1 + \gamma_2) y - \gamma_1 (y - p)), \end{aligned} \tag{8}$$

where $F(\cdot)$ is a strictly increasing CDF, the shape restrictions of Theorem 1 amount to requiring $\gamma_1 \leq 0$ and $\gamma_1 + \gamma_2 \leq 0$. While the first inequality is intuitive, and simply says that own price effect is negative, the second condition $\gamma_1 + \gamma_2 \leq 0$ is not a priori obvious, and shows the additional restriction implied by budget-constrained utility maximization. Now, applying Theorem 1, we obtain

$$\begin{aligned} & F(\gamma_0 + (\gamma_1 + \gamma_2) y - \gamma_1 (y - p)) \\ = \; & \Pr(V \leq F(\gamma_0 + (\gamma_1 + \gamma_2) y - \gamma_1 (y - p))) \\ = \; & \Pr\left(\frac{F^{-1}(V) - \gamma_0 + \gamma_1 (y - p)}{\gamma_1 + \gamma_2} \geq y \right), \end{aligned}$$

since $\gamma_1 + \gamma_2 < 0$ (note that the condition $\lim_{a \downarrow -\infty} q(a, b) = 1$ for each b rules out $\gamma_1 + \gamma_2 = 0$), implying the rationalizing utility functions

$$\begin{aligned} U_1(y - p, V) &= \frac{F^{-1}(V) - \gamma_0}{\gamma_1 + \gamma_2} + \underbrace{\left(\frac{\gamma_1}{\gamma_1 + \gamma_2} \right)}_{>0} (y - p), \\ U_0(y, V) &= y. \end{aligned}$$

where $V \simeq U(0, 1)$.

Remark 3 *Note that since the restrictions $\gamma_1 \leq 0$ and $\gamma_1 + \gamma_2 \leq 0$ are linear in parameters, it is computationally straightforward to maximize a globally concave likelihood, such as probit or logit, subject to these constraints.*

The above discussion also applies to *semiparametric* models where one need not specify the exact functional form of $F(\cdot)$. For example, the semiparametric method of Bhattacharya (2008), which only utilizes the strict monotonicity of the CDF $F(\cdot)$, can be applied to estimate the binary choice model, subject to our sign restriction and standard scale-normalization, viz. $\gamma_1 = -1$ and $\gamma_1 + \gamma_2 \leq 0$, i.e. using the specification that $\bar{q}(p, y)$ is a strictly increasing function of the linear index $-p + \gamma_2 y$ with $\gamma_2 \leq 1$.

E. Random Coefficients: An alternative parametric specification in this context is a random coefficient structure, popular in IO applications. It takes the form

$$\begin{aligned}
& \Pr\left(1|price = p, income = y\right) \\
= & \int F\left(\gamma_1 p + \gamma_2 y\right) dG\left(\gamma_1, \gamma_2, \theta\right) \\
= & \int F\left(\left(\gamma_1 + \gamma_2\right) y - \gamma_1 \left(y - p\right)\right) dG\left(\gamma_1, \gamma_2, \theta\right) \\
\equiv & H\left(y, y - p, \theta\right),
\end{aligned}$$

where γ_1 and γ_2 are now random variables with joint distribution $G(\cdot,\cdot,\theta)$, indexed by an unknown parameter vector θ, and $F(\cdot)$ is a specified CDF (e.g. a logit). Theorem 1 then implies that the distribution $G(\cdot,\cdot,\theta)$ must be such that the choice probability function $H(\cdot,\cdot,\cdot)$ satisfies $\frac{\partial}{\partial y} H(y,\cdot,\theta) \leq 0$ and $\frac{\partial}{\partial(y-p)} H(\cdot, y-p, \theta) \geq 0$. One way to guarantee this would be to specify the support of γ_1 and of $\gamma_1 + \gamma_2$ to lie in $(-\infty, 0)$. Using Theorem 1, a utility structure that would rationalize such a model is:

$$U_1\left(y - p, \eta\right) = h\left(y - p, V, \theta\right); \;\; U_0\left(y, \eta\right) = y,$$

where $V \simeq U(0,1)$, and $h(y-p, v, \theta)$ is $\sup\{x : H(x, y-p, \theta) \geq v\}$.[3]

F. Counterfactuals: Theorem 1 can be used to nonparametrically predict theory-consistent choice probabilities on counterfactual, i.e. previously unobserved, budget-sets. Obviously, without shape restrictions, there is no nonparametric restriction on demand on counterfactual budgets. To see how to use shape-restrictions, let A denote the set of (p, y) observed in the data. Then, using part (i) condition (A) of our theorem, the probability $\bar{q}(p', y')$ of buying at counterfactual (i.e. previously unobserved) price p' and income y' can be bounded as

$$\bar{q}\left(p', y'\right) \equiv q\left(y', y' - p'\right) \in \left[\sup_{\substack{(p,y)\in A:\ y \geq y', \\ y - p \leq y' - p'}} q\left(y, y - p\right), \inf_{\substack{(p,y)\in A:\ y \leq y', \\ y - p \geq y' - p'}} q\left(y, y - p\right)\right]. \tag{9}$$

The above calculation is extremely simple; for example, the lower bound requires collecting those observed budget sets (p, y) in the data that satisfy $y \geq y'$, $y - p \leq y' - p'$ (a one-line command in STATA), evaluating choice probabilities on them, and sorting these values.

[3] Note that an alternative preference distribution producing the same choice probabilities is given by $U_1(y-p,\eta) = -\gamma_1(y-p)$, $U_0(y,\eta) = \gamma_0 - (\gamma_1 + \gamma_2) y$, $\gamma_0 \perp (\gamma_1, \gamma_2)$, $\gamma_0 \simeq F(\cdot)$, $(\gamma_1, \gamma_2) \simeq G(\cdot,\cdot,\theta)$, $\gamma_1 < 0$, $\gamma_1 + \gamma_2 \leq 0$ w.p.1. This shows that the rationalizing preference distribution may not be unique.

G. Welfare bounds: Given bounds on choice probabilities, one can obtain lower and upper bounds on economically interesting functionals thereof, such as average welfare. For example, the average compensating variation – i.e. utility preserving income compensation – corresponding to a price change from p_0 to p_1 at income y is given by $\int_{p_0}^{p_1} q(y+p-p_0, y-p_0)\,dp$ (c.f. Bhattacharya, 2015).[4] This requires prediction of demand on a continuum of budget sets, viz. $\{q(y+p-p_0, y-p_0) : p \in [p_0, p_1]\}$. Now, it follows from our discussion above, and by Theorem 1, that pointwise bounds on $q(y+p-p_0, y-p_0)$ are given by

$$\begin{aligned} L(y+p-p_0, y-p_0) &\equiv \sup_{(p',y')\in A,\ y'-p'\leq y-p_0,\ y'\geq y+p-p_0} q(y', y'-p') \\ &\leq q(y+p-p_0, y-p_0) \\ &\leq \inf_{(p',y')\in A,\ y'-p'\geq y-p_0,\ y'\leq y+p-p_0} q(y', y'-p') \equiv M(y+p-p_0, y-p_0) \end{aligned} \tag{10}$$

This implies that average CV at y is bounded below by $\int_{p_0}^{p_1} L(y+p-p_0, y-p_0)\,dp$, and above by $\int_{p_0}^{p_1} M(y+p-p_0, y-p_0)\,dp$.

We can, in fact, make a stronger statement, viz., that the smallest set $S(y, p_0, p_1)$ containing *all feasible values* of the average CV, based on the restrictions of Theorem 1, is given by the interval

$$\begin{aligned} & I(y, p_0, p_1) \\ = & \left[\int_{p_0}^{p_1} L(y+p-p_0, y-p_0)\,dp, \int_{p_0}^{p_1} M(y+p-p_0, y-p_0)\,dp\right]. \end{aligned}$$

This assertion requires a justification, because $I(y, p_0, p_1)$ includes integrals of functions that violate the shape restrictions of Theorem 1 but nonetheless satisfy the pointwise bounds (10). That justification is as follows. First note that by definition, the *set* $S(y, p_0, p_1)$ is given by

$$S(y, p_0, p_1) = \left[\int_{p_0}^{p_1} f(y+p-p_0, y-p_0)\,dp,\ f \in \mathcal{F}\right],$$

where $\mathcal{F}$ is the collection of all functions $f(\cdot,\cdot) : R \times R \to [0,1]$, satisfying the conditions (i) of Theorem 1, viz. non-increasing and continuous in the first argument and non-decreasing in the second argument, satisfying $L(y+p-p_0, y-p_0) \leq f(y+p-p_0, y-p_0) \leq M(y+p-p_0, y-p_0)$ for all $p \in [p_0, p_1]$.[5] We want to show that $I(y, p_0, p_1) = S(y, p_0, p_1)$.

First, note that $S(y, p_0, p_1) \sqsubseteq I(y, p_0, p_1)$, because by definition, $L(y+p-p_0, y-p_0) \leq f(y+p-p_0, y-p_0) \leq M(y+p-p_0, y-p_0)$ for each p, and $I(y, p_0, p_1)$ is a connected interval.

[4] The results in Bhattacharya (2015) are stated in terms of the standard forms of choice probabilities, viz. $\bar{q}(p, y)$ in our notation above. In particular, average CV is $\int_{p_0}^{p_1} \bar{q}(p, y+p-p_0)\,dp$ which, in our present notation, is $\int_{p_0}^{p_1} q(y+p-p_0, y-p_0)\,dp$.

[5] The limit condition (B) in Theorem 1 can be dropped in defining S because y, p_0, p_1 are all finite.

Now, we show that $I(y,p_0,p_1) \sqsubseteq S(y,p_0,p_1)$. To see this, note that we can write any $i \in I(y,p_0,p_1)$ as

$$\begin{aligned}
i &= \lambda(i) \times \int_{p_0}^{p_1} L(y+p-p_0, y-p_0)\, dp + (1-\lambda(i)) \times \int_{p_0}^{p_1} M(y+p-p_0, y-p_0)\, dp \\
&= \int_{p_0}^{p_1} \left[\underbrace{\lambda(i) \times L(y+p-p_0, y-p_0) + (1-\lambda(i)) \times M(y+p-p_0, y-p_0)}_{=H^{\lambda(i)}(y+p-p_0, y-p_0),\text{ say}} \right] dp,
\end{aligned}$$

for some real number $\lambda(i) \in (0,1)$. But by definition, for every $\lambda \in [0,1]$, the function

$$\begin{aligned}
& H^{\lambda}(y+p-p_0, y-p_0) \\
\equiv\; & \lambda \times L(y+p-p_0, y-p_0)\, dp + (1-\lambda) \times M(y+p-p_0, y-p_0)
\end{aligned}$$

belongs to $\mathcal{F}$, since both $L(y+p-p_0, y-p_0)$ and $M(y+p-p_0, y-p_0)$, by definition, belong to $\mathcal{F}$, and monotonicity and continuity are preserved under convex additions. Hence the integral $i = \int_{p_0}^{p_1} H^{\lambda(i)}(y+p-p_0, y-p_0)\, dp \in S(y,p_0,p_1)$, and thus $I(y,p_0,p_1) \sqsubseteq S(y,p_0,p_1)$. Intuitively, even if $I(y,p_0,p_1)$ contains the integral (say of value v) of a function satisfying the pointwise bounds but not the shape restrictions, there is another function satisfying the shape restrictions and respecting the same pointwise bounds, *whose integral* has the same magnitude v.

H. Compatibility with SRP: The welfare calculation above requires prediction of demand on a continuum of budget sets indexed by $p \in [p_0, p_1]$, which is operationally difficult – if not practically impossible – to implement, using the finite-dimensional matrix equation based SRP approach. But in simple cases where there are a small, countably finite number of budget sets, and it is easy to verify the SRP conditions, a natural question is whether our shape restrictions i(A) of Theorem 1 are compatible with the SRP based criterion for rationalizability; condition i(B) and i(C) of Theorem 1 are of course irrelevant in such cases. Indeed, it is not hard to show that our shape restrictions are in fact *necessary* for the SRP criterion to be satisfied. To see this, suppose we observe behavior on two budget sets corresponding to price and income equal to (p^1, y) and (p^2, y). Let $a_0 = y$ and $a_1^j = y - p^j$ for $j = 1, 2$. Then there are three alternatives to consider, viz. $(0, a_0)$, $(1, a_1^1)$ and $(1, a_1^2)$. WLOG assume $p^1 < p^2$, i.e. $a_1^1 > a_1^2$. Under nonsatiation in numeraire, there are 3 possible preference profiles in the population, given by (i) $(0, a_0) \succ (1, a_1^1) \succ (1, a_1^2)$, (ii) $(1, a_1^1) \succ (0, a_0) \succ (1, a_1^2)$ and (iii) $(1, a_1^1) \succ (1, a_1^2) \succ (0, a_0)$; assume the population proportions of these three profiles are (π_1, π_2, π_3), respectively. Let $q(a_0, a_1^1)$, $q(a_0, a_1^2)$ denote choice probabilities of alternative 1 on the two budgets, respectively. Then the SRP approach asks whether matrix equation

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} q\left(a_0, a_1^1\right) \\ q\left(a_0, a_1^2\right) \end{bmatrix}, \text{ i.e.}$$

$$\pi_2 + \pi_3 = q\left(a_0, a_1^1\right), \ \pi_3 = q\left(a_0, a_1^2\right), \tag{11}$$

has a solution (π_1, π_2, π_3) in the unit positive simplex. Clearly, we need that $\pi_2 + \pi_3 \geq \pi_3$ (guaranteeing $\pi_2 \geq 0$) which is precisely our shape restriction $q\left(a_0, a_1^1\right) \geq q\left(a_0, a_1^2\right)$ (as $a_1^1 > a_1^2$).

Similarly, by considering the budget sets $\left(p^1, y^1\right)$ and $\left(p^2, y^2\right)$ with $y^1 < y^2$ and $y^1 - p^1 = y^2 - p^2 \equiv a_1$, say and $a_0^1 \equiv y^1 < y^2 \equiv a_0^2$, one can show that our shape restriction $q\left(a_0^1, a_1\right) \geq q\left(a_0^2, a_1\right)$ (as $a_0^1 < a_0^2$) is necessary for the SRP condition analogous to (11) to have an admissible solution. With more budget sets, the corresponding higher dimensional matrix equations analogous to (11) quickly become operationally impractical and cumbersome, as is well-known in the literature (see introduction). In contrast, our shape-restrictions, by being global conditions on the $q(\cdot, \cdot)$ functions, remain invariant to which and how many budget sets are considered. Furthermore, we already know via Theorem 1 above, that these shape restrictions are also *sufficient* for rationalizability for *any* collection – finite or infinite – of budget sets.[6]

I. Observed Covariates: One can accommodate observed covariates in our theorem. For example, let X denote a vector of observed covariates, and let $q(a, b, x)$ denote the choice probability when $Y = a$, $Y - P = b$ and $X = x$. If for each fixed x, $q(a, b, x)$ satisfies the same properties as (i) A-C in the statement of Theorem 1, then letting

$$q^{-1}(u, b, x) \stackrel{def}{=} \sup\{z : q(z, b, x) \geq u\},$$

we can rationalize the choice probabilities $q(a, b, x)$ by setting $W_1(y - p, V, x) \equiv q^{-1}(V, y - p, x)$ and $W_0(y, V, x) \equiv y$, where $V \simeq U(0, 1)$.

J. Endogeneity: Our results in Theorem 1 are stated in terms of *structural* choice probabilities $q(\cdot, \cdot)$. If budget sets are independent of unobserved heterogeneity (conditional on observed covariates), then these structural choice probabilities are equal to the observed conditional choice probabilities, i.e.,

$$q(a, b) = \Pr(1|Y = a, Y - P = b).$$

[6] It does not seem possible to show directly, i.e. *without using Theorem 1*, that our shape restrictions are also *sufficient* for existence of admissible solutions to the analog of (11) corresponding to *every* arbitrary collection of budget sets. But given theorem 1, this exercise is probably of limited interest.

To date, all existing results on rationalizability of demand under heterogeneity, including McFadden and Richter, 1990, Lewbel, 2001 and Hausman-Newey, 2016 maintain independence. If the independence condition is violated (even conditional on observed covariates), then Theorem 1 continues to remain valid as stated, since it concerns the structural choice probability $q(\cdot,\cdot)$, but consistent estimation of $q(\cdot,\cdot)$ will be more involved. In applications, if endogeneity of budget sets is a potential concern, then it would be advisable to estimate structural choice-probabilities using methods for estimating average structural functions. A specific example is the method of control functions, c.f. Blundell and Powell, 2003, which requires that $\eta \perp (P,Y)|V$, where V is an estimable "control function" – typically a first stage residual from a regression of endogenous covariates on instruments.

References

1. Anderson, S.P., De Palma, A. and Thisse, J.F. 1992. Discrete choice theory of product differentiation. MIT press.

2. Bhattacharya, D., 2015. Nonparametric welfare analysis for discrete choice. Econometrica, 83(2), pp.617-649.

3. Bhattacharya, D., 2018. Empirical welfare analysis for discrete choice: Some general results. Quantitative Economics, 9(2), pp.571-615.

4. Bhattacharya, D., 2008. A Permutation-Based Estimator for Monotone Index Models. Econometric Theory, 24(3), pp.795-807.

5. Blundell, R., and James L. Powell, 2003. Endogeneity in nonparametric and semiparametric regression models. Econometric society monographs 36: 312-357.

6. Dette, H., Hoderlein, S. and Neumeyer, N., 2016. Testing multivariate economic restrictions using quantiles: the example of Slutsky negative semidefiniteness. Journal of Econometrics, 191(1), pp.129-144.

7. Hausman, J.A. and Newey, W.K., 2016. Individual heterogeneity and average welfare. Econometrica, 84(3), pp.1225-1248.

8. Kitamura, Y. and Stoye, J., 2016. Nonparametric analysis of random utility models, forthcoming, Econometrica.

9. Lee, Y.Y. and Bhattacharya, D., 2017. Applied welfare analysis for discrete choice with interval-data on income, mimeo. UC Irvine.

10. Lewbel, A., 2001. Demand Systems with and without Errors. American Economic Review, pp. 611-618.

11. McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior.

12. McFadden, D. and Richter, M.K., 1990. Stochastic rationality and revealed stochastic preference. Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO, pp.161-186.

13. McFadden, D. 2005. Revealed Stochastic Preference: A Synthesis. Economic Theory 26(2): 245–264.

14. Matzkin, R.L., 1992. Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. Econometrica, pp.239-270.

15. Train, K.E., 2009. Discrete choice methods with simulation. Cambridge University Press.