*Econometrica*, Vol. 89, No. 1 (January, 2021), 457–474

# THE EMPIRICAL CONTENT OF BINARY CHOICE MODELS

DEBOPAM BHATTACHARYA
Department of Economics, University of Cambridge

An important goal of empirical demand analysis is choice and welfare prediction on counterfactual budget sets arising from potential policy interventions. Such predictions are more credible when made without arbitrary functional-form/distributional assumptions, and instead based solely on economic rationality, that is, that choice is consistent with utility maximization by a heterogeneous population. This paper investigates nonparametric economic rationality in the empirically important context of binary choice. We show that under general unobserved heterogeneity, economic rationality 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, that is, 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/nonparametric models in practical applications, and (c) provide computationally simple, theory-consistent bounds on demand and welfare predictions on counterfactual budge sets.

KEYWORDS: Binary choice, general heterogeneity, income effect, utility maximization, integrability/rationalizability, Slutsky inequality, shape restrictions.

## 1. INTRODUCTION

MANY IMPORTANT ECONOMIC DECISIONS faced by individuals are binary in nature, including labor force participation, retirement, college enrollment, adoption of a new technology or health product, participation in a job-training program, etc. 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, that is, whether there are analogs of Slutsky restrictions for binary 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 for prediction of demand and welfare in situations involving counterfactual, that is, 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, namely 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

---

Debopam Bhattacharya: debobhatta@gmail.com

The author would like to thank the editor, three anonymous referees, Michael Floater, 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, Project 681565 is gratefully acknowledged.

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, quasilinear preferences implying absence of income effects and/or parametrically specified heterogeneity distributions (cf. 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, producing an individual-level cross-sectional dataset that records prices, individual income and the choice made by the individual.[1] 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 $y$, for example, $\bar{q}(p, y) = F(\gamma_0 + \gamma_1 p + \gamma_2 y)$, 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 nonstochastic demand functions. More recently, for the case of two continuous goods (i.e., a good of interest plus the numeraire) under general heterogeneity, Dette, Hoderlein, and Neumayer (2016) have shown that constrained utility maximization implies quantiles of demand satisfy standard Slutsky negativity, and

---

[1] As a referee has correctly commented, income plays a prominent role in this paper, unlike many existing empirical applications which ignore the role of income.

Hausman and Newey (2016) have shown that the two are in fact equivalent. 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 Dette et al. and 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 "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; cf. 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; cf. Anderson, De Palma, and Thisse (1992), page 54–55 and Kitamura and Stoye (2016), Section 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 of demand 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 the important practical problem of 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 $\bar{q}(p, y)$; that is, if each member of the entire population were offered income $y$ and price $p$, then a fraction $\bar{q}(p, y)$ would buy the good. For now, we implicitly condition our analysis on observed covariates, and later show how to incorporate them into the results. We will show that these choice probabilities will be consistent with utility maximization by a heterogeneous population if and only if the following Slutsky-like conditions[2] hold:

$$\frac{\partial}{\partial p}\bar{q}(p,y) \le 0 \quad \text{and} \quad \frac{\partial}{\partial p}\bar{q}(p,y) + \frac{\partial}{\partial y}\bar{q}(p,y) \le 0. \tag{1}$$

For establishing this result, it will be convenient to rewrite the choice probabilities in an equivalent way as $q(y, y-p) = \bar{q}(p,y)$. Indeed, one can go back and forth between the two specifications because $\bar{q}(c,d) \equiv q(d, d-c)$ and $q(a,b) \equiv \bar{q}(a-b, a)$. The $q(y, y-p)$ formulation is motivated by the fact that given the budget set $(P,Y) = (p,y)$, an individual faces choice between the bundles $(0, y)$ and $(1, y-p)$; thus $q(\cdot,\cdot)$ is an equivalent representation of choice probabilities as functions of the income left over upon choosing options 0 and 1, respectively. For ease of exposition, we will state our results in terms of $q(\cdot,\cdot)$, and show that under smoothness they reduce to restriction (1) on $\bar{q}(\cdot,\cdot)$.

The following theorem 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.

To formally state the theorem, we introduce some notation. Let $\bar{\Omega}$ denote the support of $(P,Y)$; let $\Omega_1 = \{y - p : (p,y) \in \bar{\Omega}\}$ denote the support of $Y - P$, and for any $a_1 \in \Omega_1$ let $\Omega_0(a_1) = \{y : (p,y) \in \bar{\Omega}, y - p = a_1\}$. Corresponding to the support $\bar{\Omega}$ of $(P,Y)$, denote the support of $(Y, Y-P)$ by $\Omega$, as shorthand for $\bigcup_{a_1 \in \Omega_1} \bigcup_{a_0 \in \Omega_0(a_1)} \{a_0, a_1\}$.

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

(I) *The structural choice probability function* $q(\cdot,\cdot) : \Omega \to [0,1]$ *satisfies that* (A)(i) $q(\cdot, y-p)$ *is nonincreasing, and* (ii) $q(y, \cdot)$ *is nondecreasing*; (B) $q(\cdot, y-p)$ *is continuous*; (C) *corresponding to any fixed value* $a_1 \in \Omega_1$, *there exist a small enough real number* $y_L(a_1) \in \Omega_0(a_1)$, *satisfying* $\lim_{y \searrow y_L(a_1), y-p=a_1} q(y, y-p) = 1$ *and a large enough real number* $y_H(a_1) \in \Omega_0(a_1)$, *satisfying* $\lim_{y \nearrow y_H(a_1), y-p=a_1} q(y, y-p) = 0$.

(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* $\eta$ *denotes unobserved heterogeneity, and a distribution* $G(\cdot)$ *of* $\eta$ *such that*

$$q(y, y-p) = \int 1\{W_0(y,\eta) \le W_1(y-p,\eta)\}\, dG(\eta),$$

*where* (A') *for each fixed* $\eta$, (i) $W_0(\cdot,\eta)$ *is continuous and strictly increasing, and* (ii) $W_1(\cdot,\eta)$ *is nondecreasing*; (B') *for any* $p, y \in \bar{\Omega}$, *it holds that* $\int 1\{W_1(y - p, \eta) = W_0(y,\eta)\}\, dG(\eta) = 0$; (C') *corresponding to any fixed* $a_1 \in \Omega_1$, *there exist a small enough real number* $y_L(a_1) \in \Omega_0(a_1)$ *and a large enough real number* $y_H(a_1) \in \Omega_0(a_1)$, *satisfying* $\lim_{y \searrow y_L(a_1), y-p=a_1} \Pr[W_0(y,\eta) \le W_1(y-p,\eta)] = 1$ *and* $\lim_{y \nearrow y_H(a_1), y-p=a_1} \Pr[W_0(y,\eta) \le W_1(y-p,\eta)] = 0$.

PROOF: In the Appendix. *Q.E.D.*

---

[2] Our main result does not need smoothness; we write the conditions with derivatives here to show the Slutsky-like form of the result.

The key step in the proof is showing that (I) implies (II). This is done by constructing the utility functions $W_0(y, \eta) = y$ and $W_1(y - p, \eta) = q^{-1}(V, y - p)$ with $q^{-1}(\cdot, y - p)$ denoting a suitably defined inverse of the function $q(\cdot, y - p)$ with respect to its first argument, and the random variable $\eta = V \sim \text{Uniform}(0, 1)$. Under conditions A, B, C of Theorem 1, this construction is then shown to imply that $\Pr[W_1(y - p, \eta) \geq W_0(y, \eta)] = q(y, y - p)$. The formal proof appears in the Appendix.

*Interpretation of conditions*: Intuitively, conditions (A/A') mean that having more numeraire ceteris paribus is (weakly) better for every consumer, that is, preferences are increasing in the amount of income left over after any choice. Condition (B/B')—the "no-tie" assumption—is standard in discrete choice models, and intuitively means that there is a continuum of tastes. Condition (C) adds to condition (A); it says that holding fixed the income left over upon choosing option 1, if the income left over upon choosing option 0 is, hypothetically, made small enough, then everyone, that is, all $\eta$, will choose option 1. In particular, $y \searrow y_L(a_1)$, $y - p = a_1$ means that starting from a situation with $y - p = a_1$, we are lowering $p$ and $y$ by equal amounts, keeping $y - p$, that is, the income left over upon choosing option 1, fixed at $a_1$ while $y$, the income left over upon choosing option 0, is lowered toward $y_L(a_1)$, that is, $q(\underbrace{y}_{\searrow y_L(a_1)}, \underbrace{y - p}_{\text{fixed at } a_1}) \nearrow 1$. A symmetric interpretation applies to $y_H(a_1)$. The following examples illustrate Condition C.

EXAMPLE 1—High and Low Price: Suppose 0, 1 denote respectively not buying and buying a binary good. Suppose preferences are such that at any income $y$, if price takes a high enough value $p^H$, for example, close to the highest income in the population, no one would buy the good; conversely, when price takes a low enough value $p^L$, for example, the good is free ($p^L = 0$) or there is a high enough reward $r > 0$ for choosing option 1 (i.e., $p^L = -r < 0$) as in conditional cash transfer programs for school-attendance, everyone (i.e., all $\eta$) will choose option 1. Then starting from $y - p = a_1 > 0$, raising $p$ toward $p^H$ while simultaneously increasing $y$ by equal amount keeping $y - p$, the income left upon buying, fixed at $a_1$, we have that $q(y, y - p) \equiv q(a_1 + p, a_1) \searrow q(a_1 + p^H, a_1) = 0$; similarly, letting $p \searrow p^L$ and $y \searrow a_1 + p^L$ while keeping fixed $y - p = a_1 > 0$, we have that $q(\underbrace{y}_{\searrow a_1 + p^L}, \underbrace{y - p}_{\text{fixed at } a_1}) \nearrow q(a_1 + p^L, a_1) = 1$. Thus $y_H(a_1) = a_1 + p^H$, and $y_L(a_1) = a_1 + p^L$.

EXAMPLE 2—Labor supply: Suppose 0, 1 denote not working and working, respectively, $y$ is nonlabor income (e.g., spousal earning or interest income from investment), and $p = -w$ is the negative of net wage received upon working, so that $q(y, y - p) = q(y, y + w)$. Here, it is natural to assume that if nonlabor income $y$ is zero, then an individual must work at any positive net wage $w$ for subsistence, so that $q(0, w) = 1$, and thus $y_L(a_1) = 0$ for any positive $a_1$. Similarly, if net wage is zero, then no one with positive nonlabor income will work, that is, $q(y, y) = 0$, and thus $y_H(a_1) = a_1$.

REMARK 1: Condition C/C', which simplify the proof of the theorem, can be dropped. In the Appendix, we provide an alternative version of the theorem without conditions (C/C'), but with a slightly stronger continuity requirement (B/B') and a significantly longer proof.

REMARK 2: Note that assumptions (A)–(C) place *no restriction* on income effects, including its sign.

In statement (II) in Theorem 1, 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 $\eta$ denoting unobserved heterogeneity. This notation allows for the case where different vectors of unobservables enter the two utilities, that is, 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), $\eta$ 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).

## 3. FURTHER 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, namely $\bar{q}(p, y) \equiv q(y, y - p)$ that is, $q(a_0, a_1) \equiv \bar{q}(a_0 - a_1, a_0)$. Then, under continuous differentiability, the shape restrictions (A) from Theorem 1 are equivalent to

$$\frac{\partial}{\partial p}\bar{q}(p, y) = -\frac{\partial q(a_0, a_1)}{\partial a_1}\bigg|_{a_0=y, a_1=y-p} \leq 0, \quad \text{by Thm 1, (Aii)} \tag{2}$$

$$\begin{aligned}\frac{\partial}{\partial p}\bar{q}(p, y) + \frac{\partial}{\partial y}\bar{q}(p, y) &= -\frac{\partial q(a_0, a_1)}{\partial a_1} + \frac{\partial q(a_0, a_1)}{\partial a_0} + \frac{\partial q(a_0, a_1)}{\partial a_1}\bigg|_{a_0=y, a_1=y-p} \\ &= \frac{\partial q(a_0, a_1)}{\partial a_0}\bigg|_{a_0=y, a_1=y-p} \leq 0, \quad \text{by Thm 1, (Ai)} \end{aligned} \tag{3}$$

for all $p, y$.[3] 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

$$\frac{\partial}{\partial p}q^*(p, y) + q^*(p, y)\frac{\partial}{\partial y}q^*(p, y) \leq 0 \quad \text{for all } p, y. \tag{4}$$

For a continuous good and under general unobserved heterogeneity, Dette, Hoderlein, and Neumeyer (2016) (building on earlier work of Hoderlein (2011)), and Hausman and Newey (2016) show that (4) also holds with $q^*(p, y)$ denoting any quantile of the demand distribution for fixed $(p, y)$. Thus, for binary choice with general heterogeneity, the forms of the Slutsky inequality (2) and (3) are different from the continuous choice counterpart (4).[4] In particular, the inequalities (2) and (3) are *linear* in $\bar{q}(\cdot, \cdot)$ (and $q(\cdot, \cdot)$), unlike (4), and hence easier to impose on nonparametric estimates of $q(\cdot, \cdot)$ using, say, shape-preserving sieves that guarantee that $\frac{\partial}{\partial a_1}\hat{q}(a_0, a_1) \geq 0$, and $\frac{\partial}{\partial a_0}\hat{q}(a_0, a_1) \leq 0$ for all $a_0, a_1$.

REMARK 3: It is tempting to think of (2) and (3) as (4) 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

---

[3] I am grateful to a referee for suggesting this way of showing the equivalence.

[4] Bhattacharya (2015) (see also Lee and Bhattacharya (2019)) noted that (2) (resp., (3)) is necessary for the CDF of equivalent variation (resp., compensating variation) resulting from price-changes to be nondecreasing.

$(0,1)$. 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 (2) and (3) *cannot* be rewritten as

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

and, as such, are different from the continuous choice counterpart (4).

REMARK 4: Our rationality conditions (A) take the form of simple monotonicity restrictions on the regression function $q(\cdot,\cdot)$. There are several papers in the Statistics literature on testing monotonicity of nonparametrically estimated regressions, for example, Ghosal, Sen, and Van Der Vaart (2000), Hall and Heckman (2000), Chetverikov (2012), etc. which can therefore be used here.

B. *Observational Equivalence*: The construction in our proof of (II) ⇒ (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 nonstochastic.[5] 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, namely

$$\begin{aligned} q(a_0,a_1) &= \Pr\big[W_1(a_1)+\eta_1 > W_0(a_0)+\eta_0\big] \\ &= \Pr\big[\eta_0-\eta_1 < W_1(a_1)-W_0(a_0)\big] \\ &= F_{\eta_0-\eta_1}\big[W_1(a_1)-W_0(a_0)\big]. \end{aligned} \tag{5}$$

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

$$\frac{\partial^2}{\partial a_0 \partial a_1}\ln\left[-\frac{\frac{\partial}{\partial a_1}q(a_0,a_1)}{\frac{\partial}{\partial a_0}q(a_0,a_1)}\right]$$

---

[5] For quantile demand in the continuous case, a result of similar spirit is discussed in Hausman–Newey (2016), pages 1228–1229, 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 (cf. Hausman–Newey (2016), and Bhattacharya (2015), and so are Slutsky negativity conditions, as discussed above.

$$
\begin{aligned}
&= \frac{\partial^2}{\partial a_0 \partial a_1} \ln\left(\frac{W_1'(a_1)}{W_0'(a_0)}\right), \quad \text{from (5)} \\
&= \frac{\partial^2}{\partial a_0 \partial a_1} \left[\ln\bigl(W_1'(a_1)\bigr) - \ln\bigl(W_0'(a_0)\bigr)\right] \\
&= 0,
\end{aligned}
$$

for every $a_0$ and $a_1$. This equality is obviously not true for a general smooth and strictly monotone $q(\cdot, \cdot)$ satisfying conditions (A)–(C) of Theorem 1.

REMARK 5: 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 distributed $U(0, 1)$. 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 $W_1(a_1, \eta) \overset{a.s.}{=} F^{-1}_{W_1(a_1,\eta)}(U(a_1))$, 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 (2) 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 nonsatiated 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 $\eta$ type consumer with income $y$, then for some prices $p < p'$ she buys at price $p'$ but does not buy at $p$. Therefore,

$$W_1(y - p, \eta) < W_0(y, \eta) < W_1\bigl(y - p', \eta\bigr),$$

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

$$W(x, y - px, \eta) < W\bigl(x', y - p'x', \eta\bigr),$$

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 nonzero 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, (2) and (3) 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, namely

$$\bar{q}(p, y) = F(\gamma_0 + \gamma_1 p + \gamma_2 y) = F\bigl(\gamma_0 + (\gamma_1 + \gamma_2)y - \gamma_1(y - p)\bigr), \tag{6}$$

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

$$\begin{aligned} &F\big(\gamma_0 + (\gamma_1 + \gamma_2)y - \gamma_1(y-p)\big) \\ &\quad = \Pr\big(V \le F\big(\gamma_0 + (\gamma_1 + \gamma_2)y - \gamma_1(y-p)\big)\big) \\ &\quad = \Pr\left(\frac{F^{-1}(V) - \gamma_0 + \gamma_1(y-p)}{\gamma_1 + \gamma_2} \ge y\right), \end{aligned}$$

where $V \simeq U(0,1)$,[6] implying the rationalizing utility functions

$$\begin{aligned} W_1(y-p, V) &= \frac{F^{-1}(V) - \gamma_0}{\gamma_1 + \gamma_2} + \underbrace{\left(\frac{\gamma_1}{\gamma_1 + \gamma_2}\right)}_{\ge 0}(y-p), \\ W_0(y, V) &= y. \end{aligned}$$

REMARK 6: Note that since the restrictions $\gamma_1 \le 0$ and $\gamma_1 + \gamma_2 \le 0$ 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* binary choice models (cf. Manski (1975), Han (1987), Klein and Spady (1993)) where one need not specify the exact functional form of $F(\cdot)$. For example, the methods of Cavanagh and Sherman (1998) and Bhattacharya (2008), which only utilize 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 $\gamma_1 + \gamma_2 \le 0$, that is, using the specification that $\bar{q}(p, y)$ is a strictly increasing function of the linear index $-p + \gamma_2 y$ with $\gamma_2 \le 1$.

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

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

where $\gamma_1$ and $\gamma_2$ are now random variables with joint distribution $G(\cdot, \cdot, \theta)$, indexed by an unknown parameter vector $\theta$, and $F(\cdot)$ is a specified CDF (e.g., a probit or logit). Theorem 1 then implies that the distribution $G(\cdot, \cdot, \theta)$ must be such that the choice probability

---

[6] We implicitly assume that for fixed $y - p$, the function $q(y, y-p)$ varies with $y$ somewhere on $S(y-p)$, and thus $\gamma_1 + \gamma_2 \ne 0$.

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 $\gamma_1$ and of $\gamma_1+\gamma_2$ to lie in $(-\infty,0)$. Using Theorem 1, a utility structure that would rationalize such a model is

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

where $V\simeq U(0,1)$, and $h(y-p,v,\theta)$ is $\sup\{x:H(x,y-p,\theta)\geq v\}$.[7]

It also follows from the above discussion that not every distribution of random coefficients $G(\cdot,\cdot,\theta)$ will lead to rationalizable choice-probability functions. In particular, the commonly used assumption that $(\gamma_1,\gamma_2)$ is bivariate normal (so that the support of $\gamma_1$ and of $\gamma_1+\gamma_2$ do not lie in $(-\infty,0)$), can lead to choice probability functions $H(\cdot,\cdot,\theta)$ that would violate the shape restrictions of Theorem 1, and thus are not rationalizable.[8]

F. *Observed Covariates*: One can accommodate observed covariates in our theorem. For example, let $X$ denote a vector of observed covariates, and let $\bar{q}(p,y,x)\equiv q(y,y-p,x)$ denote the choice probability when $Y=y$, $Y-P=y-p$ and $X=x$. If for each fixed $x$, $q(\cdot,\cdot,x)$ satisfies the same properties as (I) A–C in the statement of Theorem 1, then letting

$$q^{-1}(u,y-p,x)\overset{\text{def}}{=}\sup\{z:q(z,y-p,x)\geq u\},$$

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

G. *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, that is,

$$q(y,y-p)=\Pr(1|Y=y,Y-P=y-p).$$

Early results on rationalizability of demand under heterogeneity, including McFadden and Richter (1990) and Lewbel (2001) worked under such 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.

---

[7] 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.

[8] As a numerical illustration, consider a random coefficient probit model

$$\Pr(1|\text{price}=p,\text{income}=y)=\int\Phi(\gamma_1 p+\gamma_2 y)\,dF(\gamma_1,\gamma_2,\theta),$$

where $\gamma_1\sim N(-1,0.1^2)$, $\gamma_2\sim N(3,0.2^2)$ and $\gamma_1\perp\gamma_2$, implying each of the probabilities of $\gamma_1\leq 0$ and $\gamma_2\geq 0$ exceeds 0.9999. Yet it can be verified numerically that, for example,

$$\begin{aligned}&\frac{\partial}{\partial p}\bar{q}(p,y)+\frac{\partial}{\partial y}\bar{q}(p,y)|_{p=1,y=1.2}\\&\quad=E[(\gamma_1+\gamma_2)\times\phi(\gamma_1+1.2\times\gamma_2)]\simeq 0.03>0.\end{aligned}$$

A specific example is the method of control functions (cf. Blundell and Powell (2003, 2004) and Imbens and Newey (2009), which require 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. The structural choice probability function can then be recovered (under regularity conditions) as the integral of the conditional choice probability given $p, y$ and realizations $v$ of the control variable $V$ over the marginal distribution of $V$. Hoderlein (2011), Hoderlein and Stoye (2014), Hausman and Newey (2016), and Kitamura and Stoye (2016) have previously discussed using control functions to estimate demand nonparametrically.

## 4. EMPIRICAL IMPLICATIONS

A practical implication of Theorem 1 is that it can be used to bound predicted choice probabilities on counterfactual, that is, previously unobserved, budget sets, for example, those arising from a potential policy intervention. Such predictions are more reliable when made nonparametrically, that is, without arbitrary functional-form/distributional assumptions on unobservables, and instead based solely on economic rationality. We now show how to obtain these nonparametric bounds using Theorem 1.

*Counterfactual Demand Bounds*: Let $\Omega$ denote the domain of definition of $\bar{q}(\cdot, \cdot)$. Let $A = \{(p^j, y^j), j = 1, \dots, N\} \sqsubset \Omega$ denote the set of $(p, y)$ observed in the data, with corresponding choice probabilities $\{\bar{q}^j, j = 1, \dots, N\} = \{q(y^j, y^j - p^j), (p^j, y^j) \in A\}$, satisfying condition (A) of our theorem. Suppose we are required to predict the probability $\bar{q}(p', y')$ of buying at a counterfactual (i.e., previously unobserved) price $p'$ and income $y'$ with $(p', y') \in \Omega \setminus A$. Then Theorem 1 implies the following bounds on this choice probability:

$$\bar{L}(p', y') = \begin{cases} \displaystyle\sup_{(p,y)\in A: y\geq y', y-p\leq y'-p'} \bar{q}(p, y) \\ \quad \text{if } \{(p, y) \in A : y \geq y', y - p \leq y' - p'\} \neq \phi, \\ 0 \quad \text{if } \{(p, y) \in A : y \geq y', y - p \leq y' - p'\} = \phi, \end{cases} \tag{7}$$

$$\bar{U}(p', y') = \begin{cases} \displaystyle\inf_{(p,y)\in A: y\leq y', y-p\geq y'-p'} \bar{q}(p, y) \\ \quad \text{if } \{(p, y) \in A : y \leq y', y - p \geq y' - p'\} \neq \phi, \\ 1 \quad \text{if } \{(p, y) \in A : y \leq y', y - p \geq y' - p'\} = \phi. \end{cases} \tag{8}$$

The above calculation is extremely simple; for example, the lower bound $\bar{L}(p', y')$ 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.

Note also that for all $(p, y) \in A$, we have that $\bar{L}(p, y) = \bar{q}(p, y) = \bar{U}(p, y)$.

PROPOSITION 1: *The bounds* (7) *and* (8) *are sharp.*

PROOF: Define $W = \{(y, y - p) : (p, y) \in A\} \cup (y', y' - p')$. Set $\bar{q}(p', y') \equiv q(y', y' - p') = c$ for any $c$ belonging to the interval defined by the bounds in (7) and (8). Then the elements of the set $\{q(y, y - p) : (p, y) \in A \cup (p', y')\}$ satisfy the shape restrictions (A) of Theorem 1 on $W$. In particular, if $(p, y) \in A$ satisfies $y > y'$, $y - p = y' - p'$, then

$$\bar{q}(p, y) \equiv q(y, y - p)$$

$$
\begin{aligned}
&\le \sup_{(\tilde{p},\tilde{y})\in A:\tilde{y}\ge y',\tilde{y}-\tilde{p}\le y'-p'} q(\tilde{y},\tilde{y}-\tilde{p}),\\
&\quad \text{since } q(\cdot,\cdot) \text{ satisfies condition (A) of Theorem 1 on } A\\
&\equiv \sup_{(\tilde{p},\tilde{y})\in A:\tilde{y}\ge y',\tilde{y}-\tilde{p}\le y'-p'} \bar{q}(\tilde{p},\tilde{y})\\
&\le c=\bar{q}(p',y');
\end{aligned}
$$

on the other hand, if $(p,y)\in A$ satisfies $y=y'$, $y-p>y'-p'$, then

$$
\begin{aligned}
\bar{q}(p,y) &\equiv q(y,y-p)\\
&\ge \inf_{(\tilde{p},\tilde{y})\in A:\tilde{y}\le y',\tilde{y}-\tilde{p}\ge y'-p'} q(\tilde{y},\tilde{y}-\tilde{p}),\\
&\quad \text{since } q(\cdot,\cdot) \text{ satisfies cond (A) of Thm 1 on } A\\
&\equiv \inf_{(\tilde{p},\tilde{y})\in A:\tilde{y}\le y',\tilde{y}-\tilde{p}\ge y'-p'} \bar{q}(\tilde{p},\tilde{y})\ge c=\bar{q}(p',y').
\end{aligned}
$$

Next, note that conditions (B) and (C) of our theorem have no empirical content, namely the countably finite set of values $\{\bar{q}^j, j=1,\dots,N\}\cup\{c\}$, in that there are no set of values $\{\bar{q}^j, j=1,\dots,N\}\cup\{c\}$ which can imply a violation of conditions (B) and (C). Therefore, the choice probabilities $\{\bar{q}^j, j=1,\dots,N\}\cup\{c\}$ corresponding to $A\cup(p',y')$ are compatible with a choice probability function $q(\cdot,\cdot)$ on a domain $G$ containing $W\cup(y',y'-p')$ and satisfying conditions (A)–(C) of Theorem 1 (for an explicit construction of such a function, see discussion on discrete support of $(P,Y)$ in the paragraph preceding Theorem 1 above). Therefore, applying Theorem 1, we conclude that there exist utility functions $W_1(a_1,V)$ and $W_0(a_0,V)=a_0$ with $V\simeq U(0,1)$ that satisfy the restrictions (A')-(C') of Theorem 1, and $\Pr[W_1(a_1,V)\ge a_0]=q(a_0,a_1)$ for all $(a_0,a_1)\in G$; in particular,

$$
\begin{aligned}
&\Pr\bigl[W_1(y^j-p^j,V)\ge y^j\bigr]=q^j,\quad j=1,\dots,N, \text{ and}\\
&\Pr\bigl[W_1(y'-p',V)\ge y'\bigr]=c. \qquad \textit{Q.E.D.}
\end{aligned}
$$

*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, that is, utility preserving income compensation—corresponding to a price increase from $p_0$ to $p_1$ at income $y$ is given by $\int_{p_0}^{p_1}\bar{q}(p,y+p-p_0)\,dp$ (cf. Bhattacharya (2015)). This requires prediction of demand on a continuum of budget sets, viz. $\{\bar{q}(p,y+p-p_0): p\in[p_0,p_1]\}$. Now, it follows from our discussion immediately above, and by Theorem 1, that pointwise sharp bounds on $\bar{q}(p,y+p-p_0)$ are given by

$$
\begin{aligned}
&\bar{L}(p,y+p-p_0)\\
&\equiv \begin{cases}
\displaystyle\sup_{(\tilde{p},\tilde{y})\in A,\tilde{y}-\tilde{p}\le y-p_0,\tilde{y}\ge y+p-p_0} \bar{q}(\tilde{p},\tilde{y})\\
\quad \text{if } \bigl\{(\tilde{p},\tilde{y})\in A,\tilde{y}-\tilde{p}\le y-p_0,\tilde{y}\ge y+p-p_0\bigr\}\ne\phi,\\
0 \quad \text{if } \bigl\{(\tilde{p},\tilde{y})\in A,\tilde{y}-\tilde{p}\le y-p_0,\tilde{y}\ge y+p-p_0\bigr\}=\phi,
\end{cases}\\
&\le \bar{q}(p,y+p-p_0)
\end{aligned}
$$

$$\begin{aligned} &\leq \begin{cases} \inf\limits_{(\tilde{p},\tilde{y})\in A, \tilde{y}-\tilde{p}\geq y-p_0, \tilde{y}\leq y+p-p_0} \bar{q}(\tilde{p},\tilde{y}) \\ \quad \text{if } \left\{(\tilde{p},\tilde{y})\in A, \tilde{y}-\tilde{p}\geq y-p_0, \tilde{y}\leq y+p-p_0\right\} \neq \phi, \\ 1 \quad \text{if } \left\{(\tilde{p},\tilde{y})\in A, \tilde{y}-\tilde{p}\geq y-p_0, \tilde{y}\leq y+p-p_0\right\} = \phi, \end{cases} \\ &\equiv \bar{M}(p, y+p-p_0). \end{aligned} \tag{9}$$

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

As for sharpness, let $L(y, y-p) = \bar{L}(p, y)$ be defined analogous to $q(y, y-p) = \bar{q}(p, y)$ above. Then the lower bound on average CV becomes $\int_{p_0}^{p_1} L(y+p-p_0, y-p_0)$. Now, by definition,

$$\begin{aligned} &L(a_0, a_1) \\ &\quad = \begin{cases} \sup\left\{\bar{q}(\tilde{p},\tilde{y}) : (\tilde{p},\tilde{y})\in A, \tilde{y}-\tilde{p}\leq a_1, \tilde{y}\geq a_0\right\} \\ \quad \text{if } \left\{(\tilde{p},\tilde{y})\in A, \tilde{y}-\tilde{p}\leq a_1, \tilde{y}\geq a_0\right\} \neq \phi, \\ 0 \quad \text{if } \left\{(\tilde{p},\tilde{y})\in A, \tilde{y}-\tilde{p}\leq a_1, \tilde{y}\geq a_0\right\} = \phi \end{cases} \end{aligned}$$

is nonincreasing in $a_0$ and nondecreasing in $a_1$, and $L(y, y-p) = q(y, y-p)$ when $(p, y) \in A$. Furthermore, for fixed value of $(y-p_0)$, as $p$ varies over the interval $[p_0, p_1]$, the function $L(y+p-p_0, y-p_0)$ can assume at most finitely many values (namely, $q(y^m, y^m-p^m)$, $m = 1, \ldots, N$) and, therefore, must necessarily be piecewise flat in $p$, with at most countably finite number of discontinuity points. Therefore, one can construct a function $Q(\cdot,\cdot)$ (see footnote below for an illustration) that (1) is continuous in the first argument, (2) equals $L(\cdot,\cdot)$ (and, therefore, $q(\cdot,\cdot)$) on $A$, (3) equals $L(\cdot,\cdot)$ everywhere else on the domain except in arbitrarily small (semiclosed) intervals around the points of discontinuity of $L(\cdot,\cdot)$, and (4) satisfies the same shape restrictions as $L(\cdot,\cdot)$; also, (5) $Q(\cdot,\cdot)$ can be trivially made to satisfy the limit conditions (C) of Theorem 1 by defining the limit points $y_L(\cdot)$, $y_H(\cdot)$ lower than the lowest and larger than the highest values respectively attained by $y$ in $A$ corresponding to any fixed value of $y-p$. Using (1), (4), and (5) and applying Theorem 1, we can rationalize $Q(\cdot,\cdot)$—which equals $q(\cdot,\cdot)$ at all the observed data points, that is, corresponding to $(p, y) \in A$—via a pair of utility functions and a uniformly distributed unobserved heterogeneity, and at the same time, $\int_{p_0}^{p_1} Q(y+p-p_0, y-p_0)\, dp$, is arbitrarily close to $\int_{p_0}^{p_1} L(y+p-p_0, y-p_0) = \int_{p_0}^{p_1} \bar{L}(p, y+p-p_0)\, dp$, since they differ only on at most finitely many intervals of arbitrarily small length. Therefore, $\int_{p_0}^{p_1} \bar{L}(p, y+p-p_0)\, dp$ is a sharp lower bound for average CV $\int_{p_0}^{p_1} q(y+p-p_0, y-p_0) \equiv \int_{p_0}^{p_1} \bar{q}(p, y+p-p_0)\, dp$.[9]

---

[9] As a simple illustration, consider a fixed $a_1 = y - p_0 \in \Omega_1$, and suppose the point $(k, a_1) \in A$, and $l < k < u$ for some real numbers $l, u$ belonging to the interval $[y, y+p_1-p_0]$ where the first argument of $L(y+p-p_0, y-p_0)$ takes its values as $p$ varies over $[p_0, p_1]$. Now suppose the lower bound function $L(\cdot.\cdot)$ satisfies

$$L(a_0, a_1) = \begin{cases} q(k, a_1) & \text{if } l \leq a_0 \leq k, \\ L(k^+, a_1) & \text{if } k < a_0 \leq u, \end{cases}$$

with $L(k^+, a_1) < q(k, a_1)$. That is, $L(\cdot,\cdot)$ equals $q(\cdot,\cdot)$ at the point $(k, a_1)$ in $A$, is nonincreasing in the first argument and is (right) discontinuous at $k$ with $L(k^+, a_1) < L(k, a_1)$. Choose $\delta \in (0, u-k)$ and define the

A symmetric line of argument implies that $\int_{p_0}^{p_1} \bar{M}(p, y + p - p_0)\, dp$ is the sharp upper bound.

## 5. CONNECTION WITH REVEALED STOCHASTIC PREFERENCE

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 (A) of Theorem 1 are compatible with the SRP based criterion for rationalizability; condition (B) and (C) of Theorem 1 are of course irrelevant in such cases. Below, we show that our shape restrictions (A) are in fact *necessary* for the SRP criterion to be satisfied.

PROPOSITION 2: *The shape restrictions* (A) *in Theorem* 1 *are necessary for McFadden–Richter's SRP conditions to hold.*

PROOF: Consider two price and income combinations $(p^1, y)$ and $(p^2, y)$. Suppose WLOG that $p^1 < p^2$, that is, $y - p^1 > y - p^2$. Let $q(y, y - p^1)$, $q(y, y - p^2)$ denote choice probabilities of alternative 1 on the two budgets, respectively. Assume, if possible, that out shape restriction A(ii) is violated, so that $q(y, y - p^1) < q(y, y - p^2)$. We will show that this implies violation of McFadden–Richter's SRP condition. Toward that end, consider three bundles $(0, y), (1, y - p^1)$ and $(1, y - p^2)$. Under nonsatiation in numeraire, there are three possible preference profiles in the population, given by (i) $(0, y) \succ (1, y - p^1) \succ (1, y - p^2)$, (ii) $(1, y - p^1) \succ (0, y) \succ (1, y - p^2)$, and (iii) $(1, y - p^1) \succ (1, y - p^2) \succ (0, y)$; assume the population proportions of these three profiles are $(\pi_1, \pi_2, \pi_3)$, respectively. Then McFadden–Richter's SRP condition is that the matrix equation

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} q(y, y - p^1) \\ q(y, y - p^2) \end{bmatrix}, \quad \text{that is,}$$

$$\pi_2 + \pi_3 = q(y, y - p^1), \quad \pi_3 = q(y, y - p^2), \tag{10}$$

has a solution $(\pi_1, \pi_2, \pi_3)$ in the unit positive simplex. But if our hypothesis holds, that is, $q(y, y - p^1) < q(y, y - p^2)$, then (10) implies $\pi_2 + \pi_3 < \pi_3$, that is, $\pi_2 < 0$, a violation.

---

function $Q(\cdot, a_1)$ as

$$Q(a_0, a_1) = \begin{cases} L(k, a_1) & \text{if } l \le a_0 \le k, \\ L(k, a_1) \times \left[1 - \dfrac{a_0 - k}{\delta}\right] + L(k^+, a_1) \dfrac{a_0 - k}{\delta} & \text{if } k < a_0 \le k + \delta, \\ L(k^+, a_1) & \text{if } k + \delta < a_0 \le u. \end{cases}$$

Then (1) $Q(\cdot, a_1)$ is continuous in the first argument, since $Q(a_0, a_1) \nearrow L(k, a_1)$ as $a_0 \searrow k$, and $\searrow L(k^+, a_1)$ as $a_0 \nearrow (k + \delta)$, (2) at the point $(k, a_1) \in A$, $Q(k, a_1) = q(k, a_1) = L(k, a_1)$, (3) $Q(\cdot, a_1)$ equals $L(\cdot, a_1)$ except on the semi-open interval $(k, k + \delta]$ of length $\delta$, (4) $Q(\cdot, a_1)$ is nonincreasing, and $Q(a_0, \cdot)$ is nondecreasing since $L(\cdot, a_1)$ is nonincreasing, and $L(a_0, \cdot)$ is nondecreasing. Finally, $\int_l^u Q(a_0, a_1)\, da_0 - \int_l^u L(a_0, a_1)\, da_0$ equals the area of the triangle with base $\delta$ and height $L(k, a_1) - L(k^+, a_1)$ thus equaling $\frac{L(k, a_1) - L(k^+, a_1)}{2} \delta$, which can be made arbitrarily close to 0 by choosing $\delta$ arbitrarily close to 0.

Next, consider the two price and income combinations $(p^1, y^1)$ and $(p^2, y^2)$ with $y^1 < y^2$ and $y^1 - p^1 = y^2 - p^2 \equiv a_1$, say. Let $q(y^1, a_1)$, $q(y^2, a_1)$ denote choice probabilities of alternative 1 on the two budgets, respectively. Now suppose our shape restriction A(i) is violated, so that $q(y^1, a_1) < q(y^2, a_1)$. Consider the three bundles $(0, y^1)$, $(0, y^2)$, and $(1, a_1)$. Under nonsatiation, there are three possible preference profiles in the population, given by (i) $(0, y^2) \succ (0, y^1) \succ (1, a_1)$, (ii) $(0, y^2) \succ (1, a_1) \succ (0, y^1)$, and (iii) $(1, a_1) \succ (0, y^2) \succ (0, y^1)$; assume the population proportions of these three profiles are $(\pi_1, \pi_2, \pi_3)$, respectively. Then SRP requires a solution $(\pi_1, \pi_2, \pi_3)$ in the unit positive simplex to

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} q(y^1, a_1) \\ q(y^2, a_1) \end{bmatrix}, \quad \text{that is,}$$

$$\pi_2 + \pi_3 = q(y^1, a_1), \qquad \pi_3 = q(y^2, a_1). \tag{11}$$

But $q(y^1, a_1) < q(y^2, a_1)$ implies that $\pi_2 + \pi_3 < \pi_3$ implying $\pi_2 < 0$, which is a violation of $(\pi_1, \pi_2, \pi_3)$ lying in the unit positive simplex. *Q.E.D.*

With more budget sets, the corresponding higher dimensional matrix equations analogous to (10) and (11) quickly become operationally impractical and cumbersome, as is well known in the literature (see the 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.[10]

## APPENDIX A: PROOF OF THEOREM 1

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

$$q(y, y - p) = \int 1\{y \le W_0^{-1}(W_1(y - p, \eta), \eta)\}\, dG(\eta)$$

whence (B') implies (B), (C') implies (C), and (A') implies (A).

We now show that (I) implies (II).

Note that (C) implies that for any $v \in [0, 1]$ and $a_1 \in \Omega_1$, the set $\{a_0 \in [y_L(a_1), y_H(a_1)] : q(a_0, a_1) \ge v\}$ is nonempty; for any fixed $a_1 \in \Omega_1$ and for $v \in [0, 1]$, define

$$q^{-1}(v, a_1) \overset{\text{def}}{=} \sup\{a_0 \in [y_L(a_1), y_H(a_1)] : q(a_0, a_1) \ge v\}, \tag{12}$$

which takes values in $[y_L(a_1), y_H(a_1)]$.[11] Also, by condition (A), $q^{-1}(v, \cdot)$ must be nondecreasing.

---

[10] It does not seem possible to show directly, that is, *without using Theorem 1*, that our shape restrictions are also *sufficient* for existence of admissible solutions to the analog of (10) and (11) corresponding to *every* arbitrary collection of budget sets. But given Theorem 1, this exercise is probably of limited interest.

[11] Here, we are implicitly assuming that $\Omega_0(a_1)$ equals (or contains) $[y_L(a_1), y_H(a_1)]$. If however the support of price and income are discrete, then $\Omega_0(a_1)$ can be a strict subset of $[y_L(a_1), y_H(a_1)]$. Then $q(\cdot, \cdot)$ is not defined at the points "in between" the points of support and, therefore, $q^{-1}(\cdot, a_1)$ in (12) is not well-defined.

Now, consider a random variable $V \simeq \text{Uniform}(0,1)$. Define $W_0(a_0, V) \overset{\text{defn}}{=} a_0$ and $W_1(a_1, V) \overset{\text{defn}}{=} q^{-1}(V, a_1)$. We will now show that $W_0(\cdot, V)$ and $W_1(\cdot, V)$ will rationalize the choice-probabilities $q(\cdot,\cdot)$, and satisfy properties (A')–(C') of our theorem.

To do so, first note that for any fixed $a_1 \in \Omega_1$, the function $1 - q(\cdot, a_1)$ is a continuous CDF by conditions A(i), B, and C of the theorem, and $q^{-1}(v, a_1)$ is, by definition, the corresponding $(1-v)$th quantile. Standard properties of quantiles (cf. Pfeiffer (1990), Section 11a, pp. 266–267), then imply the following three results (for completeness, we state and prove these results formally as a claim below this proof):

*Result* (i): for any $a_1 \in \Omega_1$ and $v \in [0,1]$, we must have that $q(q^{-1}(v, a_1), a_1) = v$ (Pfeiffer (1990), p. 267, property 6);

*Result* (ii): for any $a_1 \in \Omega_1$, $a_0 \in [y_L(a_1), y_H(a_1)]$ and $v \in [0,1]$, we have $q(a_0, a_1) \geq v \Leftrightarrow a_0 \leq q^{-1}(v, a_1)$ (Pfeiffer (1990), p. 266 property 1);

*Result* (iii): for any $a_1 \in \Omega_1$, the function $q^{-1}(\cdot, a_1)$ is one-to-one on [0, 1] (Consequence of *Result* (i)).

Now, for $V \simeq \text{Uniform}(0,1)$, it follows from *Result* (ii) that

$$\Pr\big(q^{-1}(V, a_1) \geq a_0\big) = \Pr\big(V \leq q(a_0, a_1)\big) = q(a_0, a_1). \tag{13}$$

Therefore, the utility functions $W_0(a_0, V) \equiv a_0$ and $W_1(a_1, V) \equiv q^{-1}(V, a_1)$ with heterogeneity $V \simeq \text{Uniform}(0,1)$ rationalize the choice probabilities $q(\cdot,\cdot)$, and satisfy all the properties specified in panel (II) of Theorem 1. In particular, $W_1(a_1, \eta)$ is nondecreasing in $a_1$ (see right after equation (12)), so (A'ii) holds; $W_0(a_0, \eta) = a_0$ trivially satisfies (A'i). Next, for $v, v' \in [0,1]$ with $v \neq v'$, we cannot have that $q^{-1}(v, a_1) = q^{-1}(v', a_1)$ by *Result* (iii); therefore,

$$\Pr\big[q^{-1}(V, a_1) = a_0\big] = 0 \quad \text{for all } a_0, \tag{14}$$

which implies property (B'). Finally,

$$\begin{aligned}
&\lim_{y \searrow y_L(a_1), y-p=a_1} \Pr\big[q^{-1}(V, y-p) \geq y\big] \\
&\overset{\text{by (13)}}{=} \lim_{y \searrow y_L(a_1), y-p=a_1} \Pr\big[q(y, y-p) \geq V\big] \\
&= \lim_{y \searrow y_L(a_1), y-p=a_1} q(y, y-p), \quad \text{since } V \simeq U(0,1) \\
&\overset{\text{by Condition (C)}}{=} 1.
\end{aligned}$$

By an analogous argument, $\lim_{y \nearrow y_H(a_1), y-p=a_1} \Pr[q^{-1}(V, y-p) \geq y] = 0$, thus satisfying (C'). *Q.E.D.*

## APPENDIX B: PROOF OF RESULTS (i), (ii), and (iii) in Theorem 1

CLAIM: *Suppose* $q(\cdot,\cdot) : \Omega \to [0,1]$ *satisfies conditions* (A), (B), (C) *of Theorem* 1, *and* $q^{-1}(\cdot,\cdot)$ *is as defined in* (12). *Then* (i) *for any* $a_1 \in \Omega_1$ *and* $v \in [0,1]$, *we must have that*

---

To cover this case, one can extend $q(\cdot,\cdot)$ to a continuous function $q^c(\cdot,\cdot)$ defined on a rectangle $\Omega^c$ containing $\Omega$ such that (i) $q^c(\cdot,\cdot)$ equals $q(\cdot,\cdot)$ on $\Omega$, (ii) $q^c(\cdot,\cdot)$ satisfies the same shape restrictions on $\Omega^c$ that are satisfied by $q(\cdot,\cdot)$ on $\Omega$, and (iii) $q^c(\cdot,\cdot)$ satisfies the limit conditions C of Theorem 1. In the Appendix in the Online Supplementary Material (Bhattacharya (2021)), we provide an explicit construction of such a function. The proof of Theorem 1 then holds with $\Omega$, $\Omega_0(\cdot)$ and $q(\cdot,\cdot)$ equaling their corresponding extensions in the case where $(P, Y)$ have discrete support.

$q(q^{-1}(v, a_1), a_1) = v$; (ii) *for any* $v \in [0, 1]$, *and any* $(a_0, a_1) \in \Omega$, *we have that* $q(a_0, a_1) \geq v \Longleftrightarrow a_0 \leq q^{-1}(v, a_1)$; (iii) *for any* $a_1 \in \Omega_1$, *the function* $q^{-1}(\cdot, a_1)$ *is one-to-one on* $[0, 1]$.

PROOF: *Claim* (i): Pick $a_1 \in \Omega_1$. For $v = 0$, we cannot have that $q(q^{-1}(v, a_1), a_1) < v$, since $q(\cdot, \cdot)$ takes values in $[0, 1]$. So let $v \in (0, 1]$, and suppose if possible that $q(q^{-1}(v, a_1), a_1) < v$. Note that $q^{-1}(v, a_1) > y_L(a_1)$ because if $q^{-1}(v, a_1) = y_L(a_1)$, then $q(q^{-1}(v, a_1)) = q(y_L(a_1), a_1) = 1 \geq v$. Therefore, $q(q^{-1}(v, a_1), a_1) < v$ implies by the continuity condition (B) that there must exist $\varepsilon > 0$ such that $q(x, a_1) < v$ for all $x \in [q^{-1}(v, a_1) - \varepsilon, q^{-1}(v, a_1)]$. But by condition (A) and the definition of $q^{-1}(\cdot, a_1)$ as the supremum in (12), we must have that $q(x, a_1) \geq v$ for all $x < q^{-1}(v, a_1)$, and in particular for $x \in [q^{-1}(v, a_1) - \varepsilon, q^{-1}(v, a_1)]$, which contradicts $q(x, a_1) < v$.

Next, for $v = 1$, we cannot have that $q(q^{-1}(v, a_1), a_1) > v$, since $q(\cdot, \cdot)$ takes values in $[0, 1]$. So let $v \in [0, 1)$ and suppose $q(q^{-1}(v, a_1), a_1) > v$. Condition (B) and (C) imply via the intermediate value theorem that $\ni x \in \Omega_0(a_1)$, such that $q(x, a_1) = v$. But by hypothesis, $q(q^{-1}(v, a_1), a_1) > v = q(x, a_1)$, so (A) implies that $x > q^{-1}(v, a_1)$, which, together with $q(x, a_1) = v$, contradicts $q^{-1}(v, a_1)$ being the supremum in (12). Therefore, $q(q^{-1}(v, a_1), a_1) = v$ for all $v \in [0, 1]$, and claim (i) is proved.

*Claim* (ii): To prove claim (ii), note that for any $v \in [0, 1]$, and any $(a_0, a_1) \in \Omega$,

$$a_0 \leq q^{-1}(v, a_1) \quad \overset{\text{by (A)}}{\Longrightarrow} \quad q(a_0, a_1) \geq \underbrace{q\big(q^{-1}(v, a_1), a_1\big)}_{=v,\ \text{by Result (i)}} \quad \Longrightarrow \quad q(a_0, a_1) \geq v. \tag{15}$$

Also, by definition of $q^{-1}(\cdot, a_1)$ as the supremum in (12), we have by (A) that

$$q(a_0, a_1) \geq v \quad \Longrightarrow \quad a_0 \leq q^{-1}(v, a_1). \tag{16}$$

Therefore, from (15) and (16), we have that $q(a_0, a_1) \geq v \Longleftrightarrow a_0 \leq q^{-1}(v, a_1)$, which proves claim (ii).

*Claim* (iii): To prove claim (iii), note that for $v, v' \in [0, 1]$ with $v \neq v'$, we cannot have that $q^{-1}(v, a_1) = q^{-1}(v', a_1)$; otherwise,

$$v \overset{\text{by Claim (i)}}{=} q\big(q^{-1}(v, a_1), a_1\big) \overset{\text{by } q^{-1}(v, a_1) = q^{-1}\left(v', a_1\right)}{=} q\big(q^{-1}\big(v', a_1\big), a_1\big) \overset{\text{by Claim (i)}}{=} v',$$

contradicting $v \neq v'$. *Q.E.D.*

## REFERENCES

ANDERSON, S. P., A. DE PALMA, AND J. F. THISSE (1992): *Discrete Choice Theory of Product Differentiation*. MIT press. [459]

BHATTACHARYA, D. (2008): "A Permutation-Based Estimator for Monotone Index Models," *Econometric Theory*, 24 (3), 795–807. [465]

——— (2015): "Nonparametric Welfare Analysis for Discrete Choice," *Econometrica*, 83 (2), 617–649. [458, 462,463,468]

——— (2018): "Empirical Welfare Analysis for Discrete Choice: Some General Results," *Quantitative Economics*, 9 (2), 571–615. [458]

——— (2021): "Supplement to 'The Empirical Content of Binary Choice Models'," *Econometrica Supplemental Material*, 89, https://doi.org/10.3982/ECTA16801. [472]

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," *Econometric Society Monographs*, 36, 312–357. [467]

BLUNDELL, R. W., AND J. L. POWELL (2004): "Endogeneity in Semiparametric Binary Response Models," *The Review of Economic Studies*, 71 (3), 655–679. [467]

CAVANAGH, C., AND R. P. SHERMAN (1998): "Rank Estimators for Monotonic Index Models," *Journal of Econometrics*, 84 (2), 351–382. [465]

CHETVERIKOV, D. (2012): "Testing Regression Monotonicity in Econometric Models," *Econometric Theory*, 1–48. [463]

DETTE, H., S. HODERLEIN, AND N. NEUMEYER (2016): "Testing Multivariate Economic Restrictions Using Quantiles: The Example of Slutsky Negative Semidefiniteness," *Journal of Econometrics*, 191 (1), 129–144. [458,462]

GHOSAL, S., A. SEN, AND A. W. VAN DER VAART (2000): "Testing Monotonicity of Regression," *The Annals of Statistics*, 28 (4), 1054–1082. [463]

HALL, P., AND N. E. HECKMAN (2000): "Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions," *The Annals of Statistics*, 28 (1), 20–39. [463]

HAN, A. K. (1987): "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 35 (2-3), 303–316. [465]

HAUSMAN, J. A., AND W. K. NEWEY (2016): "Individual Heterogeneity and Average Welfare," *Econometrica*, 84 (3), 1225–1248. [459,462,463,467]

HODERLEIN, S. (2011): "How Many Consumers Are Rational?" *Journal of Econometrics*, 164 (2), 294–309. [462,467]

HODERLEIN, S., AND J. STOYE (2014): "Revealed Preferences in a Heterogeneous Population," *Review of Economics and Statistics*, 96 (2), 197–213. [467]

IMBENS, G. W., AND W. K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77 (5), 1481–1512. [467]

KITAMURA, Y., AND J. STOYE (2016): "Nonparametric Analysis of Random Utility Models," *Econometrica*, 86 (6), 1883–1909. [459,467]

KLEIN, R. W., AND R. H. SPADY (1993): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 387–421. [465]

LEE, Y. Y., AND D. BHATTACHARYA (2019): "Applied Welfare Analysis for Discrete Choice With Interval-Data on Income," *Journal of Econometrics*, 211 (2), 361–387. [462]

LEWBEL, A. (2001): "Demand Systems With and Without Errors," *American Economic Review*, 611–618. [458, 466]

MANSKI, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3 (3), 205–228. [465]

MATZKIN, R. L. (1992): "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, 239–270. [458]

MCFADDEN, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior." [458]

——— (2005): "Revealed Stochastic Preference: A Synthesis," *Economic Theory*, 26 (2), 245–264. [459]

MCFADDEN, D., AND M. K. RICHTER (1990): "Stochastic Rationality and Revealed Stochastic Preference," in *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz*. Westview Press, 161–186. [459, 466]

PFEIFFER, P. E. (1990): *Probability for Applications*. Springer Science & Business Media. [472]

TRAIN, K. E. (2009): *Discrete Choice Methods With Simulation*. Cambridge University Press. [458]

---

*Co-editor Ulrich K. Müller handled this manuscript.*

*Manuscript received 5 November, 2018; final version accepted 18 September, 2020; available online 23 September, 2020.*