Econometrica, Vol. 83, No. 2 (March, 2015), 617–649

NONPARAMETRIC WELFARE ANALYSIS FOR DISCRETE CHOICE

BY DEBOPAM BHATTACHARYA[1]

We consider empirical measurement of equivalent variation (EV) and compensating variation (CV) resulting from price change of a discrete good using individual-level data when there is unobserved heterogeneity in preferences. We show that for binary and unordered multinomial choice, the marginal distributions of EV and CV can be expressed as simple closed-form functionals of conditional choice probabilities under essentially unrestricted preference distributions. These results hold even when the distribution and dimension of unobserved heterogeneity are neither known nor identified, and utilities are neither quasilinear nor parametrically specified. The welfare distributions take simple forms that are easy to compute in applications. In particular, average EV for a price rise equals the change in average Marshallian consumer surplus and is smaller than average CV for a normal good. These nonparametric point-identification results fail for ordered choice if the unit price is identical for all alternatives, thereby providing a connection to Hausman–Newey's (2014) partial identification results for the limiting case of continuous choice.

KEYWORDS: Binary choice, multinomial choice, ordered choice, applied welfare analysis, compensating variation, equivalent variation, unobserved heterogeneity, unrestricted heterogeneity, nonparametric identification.

1. INTRODUCTION

THIS PAPER CONCERNS the empirical measurement of money-metric welfare with regard to goods that are consumed or, more generally, decisions that are made, in discrete form. The specific focus is on evaluating welfare effects of price changes, such as those brought about by taxes and subsidies. Examples include, inter alia, the effects of taxing unemployment benefits on exiting unemployment, of fare hikes on choice of mode of transportation, and of price discounts on consumers' brand choices in supermarkets. The setting is where the researcher observes realizations of the discrete decision at the individual level from microdata sets that also record the individual's characteristics, including income, and prices faced by her with regard to the discrete decision. The goal is to estimate the impact on individual welfare, measured in terms of income compensation, of a hypothetical change in price. The analysis incorporates unobserved, individual heterogeneity in utility functions, and focuses on recovering the distribution of the impact of price change on individual welfare arising from such heterogeneity without restricting the nature of heterogeneity or specifying the functional form of utilities.

[1]I am grateful to Richard Blundell, Martin Browning, Ian Crawford, Eric Gautier, Arthur Lewbel, and Whitney Newey for helpful feedback and encouragement. I would also like to thank the co-editor and four anonymous referees for their suggestions, which led to substantial improvement in both the content and the exposition. All remaining errors are mine.

 DOI: 10.3982/ECTA12574

Overview of Results

Our key results for binary and unordered multinomial choice are (i) the marginal distribution, and, consequently, the average of compensating variation (CV) and equivalent variation (EV) corresponding to a price change, are nonparametrically point-identified solely from choice probabilities; (ii) for a price rise, the average EV is identical to the change in average Marshallian consumer surplus even if utility is not quasilinear in income; (iii) the average CV exceeds the average EV if the good is normal; (iv) the above conclusions hold even if the dimension and distribution of heterogeneity are *not* identified. These results are fully nonparametric in the sense that they do not require one to specify the dimension or distribution of heterogeneity and the functional form of utilities, thus allowing for extremely general preference distributions in the population. A practical advantage of our results is that they express welfare distributions as simple, closed-form functionals of choice probabilities and thus can be easily computed in applications. Finally, we show that (v) nonparametric point identification of EV and CV distributions fails in the case of *ordered* choice with three or more alternatives if the unit price is identical across all alternatives. To our knowledge, these results constitute the first set of results in the econometric literature on the nonparametric identification of money-metric welfare distributions for discrete choice.

Related Literature

In econometrics, a large literature exists on demand estimation from individual consumption data incorporating unobserved heterogeneity (cf. Lewbel (2001) and references therein). Indeed, an important use of such demand estimates is the calculation of welfare effects of price change arising from taxes and subsidies. For demand of a good that is consumed in continuous quantities, such as gasoline, Hausman (1981), in a seminal paper, formulated the nonparametric identification and parametric estimation of exact welfare effects of a price change. Vartia (1983) provided an alternative computational approach to the same problem. Hausman and Newey (1995) extended these analyses by formulating semiparametric estimation of welfare effects and developing the corresponding theory of statistical inference. Their methods cannot be directly used in discrete choice settings where the effect of a price change on individual utilities depends, in a fundamental way, on the discreteness of choice possibilities as well as on general individual heterogeneity. Specifically, in discrete choice scenarios, corner solutions are generic and this makes it difficult to recover the compensation functions using the differential-equation-based approach of the above papers.

In the discrete choice setting, Domencich and McFadden (1975; DM75 henceforth) made the strong assumption that utility is quasilinear, that is, additively separable in income, implying that choice probabilities are unaffected by income changes. Under this simplifying but restrictive assumption, Marshallian and Hicksian welfare measures are identical and average welfare effects of price change can be expressed as the integral of the ordinary Marshallian choice probabilities (cf. DM75, pp. 94–99). In a highly influential subsequent paper, Small and Rosen (1981; SR81 henceforth) investigated the measurement of welfare effects of price and quality change for discrete choice. In their empirical formulation, SR81 introduced additive scalar heterogeneity in utility functions but assumed that the discrete good is sufficiently unimportant to the consumer so that income effects from price or quality changes are negligible (cf. SR81, p. 124, Assumptions (a) and (b)), thereby equating Marshallian and Hicksian welfare measures. More recently, Herriges and Kling (1999; HK99) and Dagsvik and Karlstrom (2005; DK05), in their analysis of the same problem, allowed utility to be nonlinear in income, and incorporated unobservables in utility but assumed that these unobservables both have a known dimension and follow a known parametric distribution for identifying and estimating the distribution of welfare effects of price changes. The HK99 and DK05 analyses also require the functional forms of utilities to be known up to finite-dimensional parameters that are either nonstochastic or stochastic with fully known distributions and with the heterogeneity entering the utility function in a known way. Apart from the usual concerns about misspecification, these parametric approaches do not clarify whether the identification arises from the functional form assumptions or is more fundamental in the sense that the choice probabilities contain all the identification-relevant information, with parametric computations being simply a convenient approximation. Indeed, we will show that our nonparametric point-identification results that hold for *unordered* multinomial choice fail for *ordered* choice—a fundamental difference that would not be apparent if one were to focus only on parametric models.

In contrast to the works cited above, the purpose of the present paper is to establish *nonparametric* point identification of the distribution of welfare effects of price change in a discrete choice setting, incorporating unobservable heterogeneity in the utility function, and assuming no knowledge of the dimension and, thus, of the distribution of these unobservables. Indeed, in many applications, it is important to allow for multiple sources of unobserved heterogeneity and it is easy to see that restricting the dimension of heterogeneity can place arbitrary restrictions on the variation of individual preferences in the population. For instance, consider the canonical nonparametric binary choice equation $Q = 1\{\beta(P, Y) + \varepsilon > 0\}$, where $\beta(P, Y)$ is an unknown function of price P and income Y, and ε is a scalar additive heterogeneity with an unknown distribution (cf. Matzkin (1992) and Manski (1975)). This model implies that if, at a specific price and income $(P, Y) = (p, y)$, an individual i prefers to buy while individual j does not (implying $\varepsilon_i > \varepsilon_j$), then at no price–income combination (p', y') can we have a situation where individual i prefers not to buy while individual j prefers to buy—an extremely strong (and untestable) restriction. This restriction is also implied by the more general model $q = 1\{\beta(P, Y, \varepsilon) > 0\}$, where ε is a scalar and $\beta(p, y, \cdot)$ is strictly increasing for all p, y. Thus allowing for heterogeneity of unrestricted dimension immediately extends the scope of the results to a much larger set of preference profiles.

The *continuous* choice analog of the present paper is Hausman and Newey (2014; HN14). They considered unobserved individual heterogeneity of unspecified dimension in utility functions and focussed on recovering average equivalent variation resulting from price change of a *continuous* consumption good, based on demand data. HN14 showed that (a) the dimension of heterogeneity is not identified from observed demand and (b) if one allows for heterogeneity of unspecified dimension, then one cannot point-identify average welfare but may obtain bounds on it. Kitamura and Stoye (2013) (see also McFadden and Richter (1991)) have also worked under heterogeneity of unspecified dimension and provided tests of consumer rationality in that framework. We are not aware of any other work that performs demand analysis at this level of generality regarding both individual heterogeneity and the form of utility functions. Hoderlein and Vanhems (2011), Blundell, Horowitz, and Perry (2012), Blundell, Kristensen, and Matzkin (2014), and Lewbel and Pendakur (2013), among others, discussed demand and welfare estimation for continuous choice under restricted heterogeneity.

For binary choice, Ichimura and Thompson (1998) and Gautier and Kitamura (2013) considered random coefficient models where the dimension of heterogeneity is specified to be equal to the number of regressors (plus one for the random intercept) and enter the individual outcome equation in a special way, namely, as scalar multipliers—one attached to each regressor. These authors provided conditions, including large support requirements for regressors, under which the distribution of these random coefficients can be nonparametrically identified up to scale normalization. Such support requirements may be restrictive in specific demand applications. Besides, estimation of the heterogeneity distribution is complicated owing to ill-posed inverse problems. The present paper shows that for the purpose of welfare analysis, this exercise is not necessary and that even the dimension of heterogeneity does not need to be specified in advance. This finding appears to be of significant practical importance because in demand applications, a key motivation for recovering the distribution of consumer heterogeneity is the calculation of welfare distributions resulting from price change.

The welfare analysis presented here is based on the “revealed preference” approach, where actual choice data are used to infer money-metric welfare impacts of hypothetical price changes. In an alternative methodology, known as the stated preference or contingent-valuation approach, a sample of individuals are asked how much money they would be willing to accept as compensation for a hypothetical policy change (e.g., a price rise). Lewbel, Linton, and McFadden (2011) have recently shown how distributions of the willingness to pay can be inferred from stated preference data without imposing parametric assumptions.

The rest of the paper is organized as follows. In Section 2.1, we analyze the leading case of binary choice. In Section 2.2, we consider the case of unordered multinomial choice. In Section 3, we discuss ordered choice. All proofs are collected in the Appendix, which also contains an example where CV and EV distributions are point-identified even when the dimension and distribution of unobserved heterogeneity are not.

2. FORMAL SETUP AND RESULTS

We first analyze the binary choice case and obtain the corresponding welfare results. We then show that the unordered multinomial case can be analyzed by reducing it to a binary case with a composite outside option and then applying the binary choice results.

2.1. *Binary Choice*

Consider an individual with income Y, who faces the choice between buying and not buying a binary good that costs P. Let W represent the quantity of numeraire that the individual consumes in addition to the binary good. Suppose that the utility realized by the individual is given by

$$\begin{cases} U_1(W,\eta) & \text{if 1 is chosen,} \\ U_0(W,\eta) & \text{if 0 is chosen.} \end{cases} \tag{1}$$

Here η is a possibly vector-valued, individual-specific taste variable *of unknown dimension*, unobserved by the econometrician, that enters the utility functions in any arbitrary way. Income and prices faced by the individual are observed by the econometrician. In addition, he may observe a set of covariates. The latter will be suppressed in the exposition below for notational clarity, that is, the entire analysis should be thought of as implicitly conditioned on these observed covariates.

Given total income Y, the budget constraint is $W + PQ = Y$, where $Q \in \{0, 1\}$ represents the binary choice. Replacing the budget constraint in (1), the consumer's realized utility is given by

$$\begin{cases} U_1(Y-P,\eta) & \text{if 1 is chosen,} \\ U_0(Y,\eta) & \text{if 0 is chosen.} \end{cases} \tag{2}$$

A consumer of type η and income Y, and facing price P chooses option 1 (buy the good) if and only if $U_1(Y - P, \eta) > U_0(Y, \eta)$.

Define the structural choice probability at hypothetical price and income (p, y) as

$$\begin{aligned} \bar{q}(p,y) &\equiv \Pr\big\{U_1(y-p,\eta) > U_0(y,\eta)\big\} \\ &\equiv \int 1\big\{U_1(y-p,\eta) > U_0(y,\eta)\big\}\, dF(\eta), \end{aligned} \tag{3}$$

where $F(\cdot)$ denotes the marginal distribution of η. This is akin to the "average structural function" defined in Blundell and Powell (2003). Our identification results in this paper are concerned with expressing the marginal distributions of welfare in terms of $\bar{q}(\cdot,\cdot)$.

When observed realizations of P and Y are jointly independent of preference heterogeneity η (conditional on observed covariates), one can obtain $\bar{q}(\cdot,\cdot)$ by a nonparametric regression of the individual's decision to buy alternative 1 on prices and income, that is, the Marshallian choice probability. Independence of preference heterogeneity and budget sets has been assumed in all preexisting research on welfare estimation for discrete choice, including DM75, SR81, and DK05. It is a maintained assumption in random coefficient models (cf. Ichimura and Thompson (1998) and Gautier and Kitamura (2013)) that are more restrictive than our setup with regard to the form and dimension of heterogeneity. Conditional independence is also assumed in Hausman and Newey (2014) in the context of continuous choice with unrestricted heterogeneity. If conditional independence fails, then $\bar{q}(\cdot,\cdot)$ can be recovered using, say, control functions (Blundell and Powell (2003)), which satisfy that η is independent of prices and income, conditional on the control functions. Petrin and Train (2010) demonstrated, both theoretically and empirically, how to use control functions to recover average structural functions in parametric multinomial choice models. Note, however, that our key results in this paper establish the relationship between welfare distributions and the structural choice probabilities $\bar{q}(\cdot,\cdot)$. These results hold regardless of whether price and income are endogenous or exogenous. The impact of endogeneity is that it affects how one could consistently estimate $\bar{q}(\cdot,\cdot)$.[2]

Now, we impose the following assumption on the utility functions—our only substantive assumption in this paper.

ASSUMPTION 1: *Suppose that for each η, $U_0(a,\eta)$ and $U_1(a,\eta)$ are continuous and strictly increasing in a. Let $U_1^{-1}(b,\eta)$ denote the unique solution x to the equation $U_1(x,\eta)=b$ and let $U_0^{-1}(b,\eta)$ denote the unique solution in x to the equation $U_0(x,\eta)=b$.*

Strict monotonicity simply says that, all else equal, more numeraire is better. The continuity condition is technical and guarantees that inverses of the utility functions are defined everywhere.

[2]Accordingly, throughout this paper, we use the term "point-identified" to mean that welfare distributions can be expressed as exact closed-form functionals of the structural choice probabilities $\bar{q}(\cdot,\cdot)$ and use the term "failure of point identification" to mean that they cannot. If $\bar{q}(\cdot,\cdot)$ itself is also point-identified from choice data, either via nonparametric regression in the exogenous case or control functions in the endogenous case, then we get a "full" identification result in that welfare distributions can be expressed directly as functions of observed data.

The above specification is far more general than the additive, scalar heterogeneity structure used in DM75 and SR81, where utility is given by

$$\begin{cases} U_1(Y-P)+\varepsilon_1 & \text{if 1 is chosen,} \\ U_0(Y)+\varepsilon_0 & \text{if 0 is chosen,} \end{cases}$$

where the functional forms of $U_1(\cdot,\cdot)$ and $U_0(\cdot,\cdot)$ are *known* (up to estimable finite-dimensional parameters), and ε_1 and ε_0 are scalar random variables that are independent of price and income and have a known distribution. DK05 (see Section 5 of their paper) considered the mixed multinomial logit-type structure

$$\text{(4)} \qquad \begin{cases} U_1(Y-P,\beta)+\varepsilon_1 & \text{if 1 is chosen,} \\ U_0(Y,\beta)+\varepsilon_0 & \text{if 0 is chosen,} \end{cases}$$

where utilities are of *known* functional form and smooth in parameters; the random coefficients β, which enter the utilities in a known way, are independent of scalar-valued additive errors ε_1, ε_0, and $(\beta, \varepsilon_1, \varepsilon_0)$ have a fully *known* joint probability distribution independent of (P, Y). DK05 derived expressions for the distribution of Hicksian welfare measures in terms of these known heterogeneity distributions and known utility functions.[3] Indeed, if (i) the dimension and distribution of unobserved heterogeneity are separately identified from choice probabilities, (ii) functional forms of utilities are known, and (iii) the key unobservables (denoted by ε's) enter as additive scalar errors in utility functions and all other unobservables (denoted by β) are either nonexistent or enter utilities in a known way and are independent of the ε's, then one can use the expressions in DK05 to calculate the distribution of Hicksian welfare measures. These conditions are restrictive and somewhat arbitrary. In the Appendix of the present paper, we provide an example of binary choice where the dimension and, thus, distribution of heterogeneity are *not* identified, and, thus, the DK05 results cannot be applied, and yet welfare distributions are nonparametrically point-identified using our results because they are based solely on the choice probability functions.[4]

The limitations of a fully parametric approach are brought out more clearly in the context of ordered choice, which is discussed in Section 3. In this case, making parametric assumptions would enable one to "calculate" the distribution functions of CV and EV exactly (see Remark 3). However, we show in Section 3 that for ordered choice, the distributions of CV and EV generically

[3]For example, DK05, p. 63, after the proof of Theorem 2, clarify that "... one can calculate the Hicksian choice probabilities readily, *provided the cumulative distribution $F^B(\cdot)$ is known....*"

[4]A referee has pointed out that even in the scalar heterogeneity case, for example, in a probit model, the variance and hence the distribution of unobserved heterogeneity is not identified from choice probabilities.

fail to be nonparametrically point-identified; thus the parametric calculations in this case would correspond to identification obtained solely via functional form assumptions. This important difference between ordered and unordered cases would not be apparent if one were only using parametrically specified utility functions and heterogeneity distributions.

In contrast, our identification results require only that utilities are continuous and strictly increasing in the numeraire (Assumption 1), and do not require specification of the functional form of the utilities, including how unobserved heterogeneity enter utilities, to impose differentiability, to specify the dimension and distribution of unobserved heterogeneity, or to assume arbitrary independence conditions among different components of heterogeneity, as in model (4) above.[5]

Welfare Analysis

Recall our setup in (2) and consider a hypothetical ceteris paribus price increase from p_0 to p_1. We wish to calculate the marginal distributions of welfare change evaluated at fixed income y that corresponds to this price change. In particular, the EV is the amount of income S to be subtracted from an η-type individual with income y and facing prices p_0 so that her (maximized) utility in this situation equals that when she was facing prices p_1 where $p_1 > p_0$. Thus $S^{\mathrm{EV}}(y, p_0, p_1, \eta)$ is the solution S to the equation

$$\begin{aligned} (5) \qquad & \max\bigl\{U_0(y-S,\eta), U_1(y-S-p_0,\eta)\bigr\} \\ & \quad = \max\bigl\{U_0(y,\eta), U_1(y-p_1,\eta)\bigr\}. \end{aligned}$$

Similarly, the CV measures the income S to be given to an individual of type η at income y and facing price $p_1 > p_0$, so that her maximized utility in this

[5]Our setup is also more general than a pure random coefficient model, which postulates that $Q = 1\{-P + \varepsilon_1 Y + \varepsilon_0 > 0\}$, where $(\varepsilon_0, \varepsilon_1)$ is a two-dimensional random vector independent of (P, Y). For example, the random coefficient model implies that for $P = p$ fixed, if an individual prefers to buy at income y and also at income $y' > y$, then she must also prefer to buy at any intermediate income $\lambda y + (1-\lambda)y'$ for $\lambda \in (0, 1)$. This is because

$$-p + \varepsilon_1\bigl(\lambda y + (1-\lambda)y'\bigr) + \varepsilon_0 = \lambda\{-p + \varepsilon_1 y + \varepsilon_0\} + (1-\lambda)\bigl\{-p + \varepsilon_1 y' + \varepsilon_0\bigr\},$$

so that if each term on the right-hand side (RHS) is positive, then so is the left-hand side (LHS). This is not imposed by our setup because

$$U_1(y-p,\eta) > U_0(y,\eta) \quad \text{and} \quad U_1\bigl(y'-p,\eta\bigr) > U_0\bigl(y',\eta\bigr)$$

need *not* imply that

$$U_1\bigl(\lambda y + (1-\lambda)y' - p, \eta\bigr) > U_0\bigl(\lambda y + (1-\lambda)y', \eta\bigr).$$

situation is equal to her maximized utility when prices were p_0 and income was y, that is, $S^{\mathrm{CV}}(y, p_0, p_1, \eta)$ is the solution S to the equation

$$\begin{aligned}(6)\qquad & \max\{U_0(y+S,\eta), U_1(y+S-p_1,\eta)\} \\ & \quad = \max\{U_0(y,\eta), U_1(y-p_0,\eta)\}.\end{aligned}$$

In what follows, we will establish the marginal distribution of $S^{\mathrm{EV}}(y, p_0, p_1, \eta)$ and $S^{\mathrm{CV}}(y, p_0, p_1, \eta)$ induced by the distribution of η. This will be done via two intermediate lemmas, which give the analytical form of individual welfare changes in terms of the underlying utility functions and unobserved heterogeneity. Using these lemmas, we will establish the key theorem, which gives the ultimate form of welfare distributions in terms of choice probabilities.

LEMMA 1: *Suppose Assumption 1 holds. Then* (i) *if* $U_1(y-p_0,\eta) \leq U_0(y,\eta)$, *then* $S^{\mathrm{EV}}(y, p_0, p_1, \eta) = 0$; (ii) *if* $U_1(y-p_1,\eta) \leq U_0(y,\eta) < U_1(y-p_0,\eta)$, *then* $S^{\mathrm{EV}}(y, p_0, p_1, \eta) = y - p_0 - U_1^{-1}(U_0(y,\eta),\eta)$; (iii) *if* $U_1(y-p_1,\eta) > U_0(y,\eta)$, *then* $S^{\mathrm{EV}}(y, p_0, p_1, \eta) = p_1 - p_0$.

These three cases correspond, respectively, to η's who (i) do not buy at the lower price, (ii) those who switch from buying at lower price to not buying at the higher price, and (iii) those who buy at both low and high price. While the zero EV is obvious for the first group, the other cases are not entirely obvious because one needs to understand how buying is affected when income is deducted from a situation of low price.

LEMMA 2: *Suppose Assumption 1 holds. Then* (i) *if* $U_1(y-p_0,\eta) \leq U_0(y,\eta)$, *then* $S^{\mathrm{CV}}(y, p_0, p_1, \eta) = 0$; (ii) *if* $U_0(y,\eta) < U_1(y-p_0,\eta) \leq U_0(y+p_1-p_0,\eta)$, *then* $S^{\mathrm{CV}}(y, p_0, p_1, \eta) = U_0^{-1}(U_1(y-p_0,\eta),\eta) - y$; (iii) *if* $U_1(y-p_0,\eta) > U_0(y+p_1-p_0,\eta)$, *then* $S^{\mathrm{CV}}(y, p_0, p_1, \eta) = p_1 - p_0$.

These three cases correspond, respectively, to η's who (i) do not buy at the lower price, (ii) those who switch from buying at lower price to not buying at the higher price *but when compensated by the amount of price change would prefer not to buy*, and (iii) switchers who when compensated by the amount of price change would prefer to buy as well as those who buy at both low and high price. While the zero CV is obvious for the first group, the other cases are not obvious because one needs to understand how buying is affected when income is raised from a situation of high price.

We now state our main result, which expresses the marginal distribution of individual CV and EV in terms of choice probabilities.

THEOREM 1: *Suppose Assumption 1 holds. Consider a price rise from* p_0 *to* p_1. *Then the EV and CV evaluated at hypothetical income* y *have marginal distributions given by*

$$\Pr\{S^{\mathrm{EV}}(y, p_0, p_1, \eta) \le a\} = \begin{cases} 0 & \text{if } a < 0, \\ 1 - \bar{q}(p_0 + a, y) & \text{if } 0 \le a < p_1 - p_0, \\ 1 & \text{if } a \ge p_1 - p_0, \end{cases} \tag{7}$$

$$\Pr\{S^{\mathrm{CV}}(y, p_0, p_1, \eta) \le a\} = \begin{cases} 0 & \text{if } a < 0, \\ 1 - \bar{q}(p_0 + a, y + a) & \text{if } 0 \le a < p_1 - p_0, \\ 1 & \text{if } a \ge p_1 - p_0, \end{cases} \tag{8}$$

where $\bar{q}(\cdot, \cdot)$ *is defined above in* (3).

Some Intuition

An intuitive interpretation of the result is as follows. First consider EV. Consider an η-type individual whose reservation price for the binary good at income y is $p_0 + t(y, \eta)$, where $0 < t(y, \eta) < p_1 - p_0$. This means that she is indifferent between buying and not buying at price $p_0 + t(y, \eta)$ when she has income y so that

$$U_0(y, \eta) = U_1\big(y - \big(p_0 + t(y, \eta)\big), \eta\big). \tag{9}$$

At any higher price, the RHS is smaller and she does not buy; at any lower price, the RHS is larger and she buys. But since $y - (p_0 + t(y, \eta)) = (y - t(y, \eta)) - p_0$, we get that

$$U_0(y, \eta) = U_1\big(\big(y - t(y, \eta)\big) - p_0, \eta\big),$$

which means that if we take away an amount of the numeraire equal to $t(y, \eta)$, then she would reach the same level of utility from buying at price p_0 as she would when not buying. Recall that since $t(y, \eta) < p_1 - p_0$, she was not buying at the higher price p_1 and getting utility $U_0(y, \eta)$, which is precisely the reference utility for calculation of her EV. The previous display therefore implies that the EV for a price increase from p_0 to p_1 is $t(y, \eta)$ for this consumer. That is, the EV equals the difference between the reservation price and the initial lower price p_0. Therefore, the probability that EV is less than a equals the proportion of individuals with reservation price less than $p_0 + a$, which, by definition, equals the fraction of individuals who do not buy at prices higher than $p_0 + a$ and is thus given by $1 - \bar{q}(p_0 + a, y)$.

The intuitive interpretation of the CV distributional result, namely (8), is slightly more involved and is as follows. Consider an η-type individual, who faces initial price p_0 and has income y. Now consider a situation where price goes up to $p_0 + a$, where $0 \leq a < p_1 - p_0$ and her income is compensated by the amount of price rise a. Suppose that in this new situation, her utility from not buying alternative 1 exceeds her utility from buying it, that is,

$$U_0(y + a, \eta) > U_1\big((y + a) - (p_0 + a), \eta\big) = U_1(y - p_0, \eta).$$

Since $U_0(\cdot, \eta)$ is strictly increasing, it is also true that $U_0(y + a, \eta) > U_0(y, \eta)$. Putting the two together yields

$$U_0(y + a, \eta) > \max\big\{U_1(y - p_0, \eta), U_0(y, \eta)\big\}.$$

For CV calculation, the RHS utility is the reference utility (see (6)) and, therefore, any compensation exceeding a will lead such a person to not buy alternative 1 and enjoy a utility level exceeding the reference utility level. Thus for such a person, the CV must not exceed a. Such individuals can be identified in the data as those who do not buy alternative 1 when they have income $y + a$ and price is $p_0 + a$, and, thus, $\Pr(S^{\mathrm{CV}} \leq a) = 1 - \bar{q}(p_0 + a, y + a)$. In particular, the CV must not exceed a for those who had switched initially from buying alternative 1 to not buying it due to the price rise from p_0 to $p_0 + a$ but who would nonetheless, upon getting compensated income $y + a$, strictly prefer to not buy alternative 1 when the price is $p_0 + a$. These individuals can be made as happy as they originally were by paying a compensation strictly less than a and letting them leave without buying alternative 1.

REMARK 1: It is important to note that the proofs of Lemmas 1 and 2, and Theorem 1 do not rely on whether the distribution of unobserved heterogeneity η is known or identified. In the last section of the Appendix, we provide an example where even the dimension of heterogeneity is not identified and yet Assumption 1 is satisfied, so that the marginal distributions of welfare changes are point-identified.

Nondecreasing Cumulative Distribution Function (CDF)

So that the CDF of CV in (8) is weakly increasing, one needs to check whether for all $a \geq 0$, the function $\bar{q}(p_0 + a, y + a)$ is nonincreasing in a for all y. Similarly, for the CDF of EV in (7) to be weakly increasing, one needs to check that for all $a \geq 0$, the function $\bar{q}(p_0 + a, y)$ is nonincreasing for all y.

Note that according to the random utility model considered here,

$$\begin{aligned}&\bar{q}(p_0+a, y+a)\\&\quad\equiv\int 1\{U_1(y+a-(p_0+a),\eta)\geq U_0(y+a,\eta)\}\,dF(\eta)\\&\quad=\int 1\{U_1(y-p_0,\eta)\geq U_0(y+a,\eta)\}\,dF(\eta),\end{aligned}$$

which is nonincreasing in a since $U_0(\cdot,\eta)$ is strictly increasing by Assumption 1. Similarly,

$$\bar{q}(p_0+a, y)=\int 1\{U_1(y-p_0-a,\eta)\geq U_0(y,\eta)\}\,dF(\eta),$$

which is nonincreasing in a since $U_1(\cdot,\eta)$ is assumed to be strictly increasing. One may test these conditions after estimating an unrestricted $\bar{q}(\cdot, y)$ or impose these restrictions during estimation.

The distributions obtained in Theorem 1 imply the following expressions for average welfare change.

COROLLARY 1: *Suppose Assumption 1 holds. Then for a price increase from p_0 to p_1, the average EV and CV evaluated at income y are given by*

$$\mu^{\mathrm{EV}}(y, p_0, p_1)=\int_{p_0}^{p_1}\bar{q}(p, y)\,dp, \tag{10}$$

$$\mu^{\mathrm{CV}}(y, p_0, p_1)=\int_{p_0}^{p_1}\bar{q}(p, y+p-p_0)\,dp, \tag{11}$$

where $\bar{q}(p, y)$ is defined in (3).

Since $\bar{q}(p, y)$ is simply the average Marshallian choice probability of alternative 1, (10) shows that the change in Marshallian consumer surplus and the average Hicksian equivalent variations are equal. This result obtains although the utility functions were not specified to be quasilinear. Indeed, for quasilinear utility, *both CV and EV* are equal to the change in Marshallian consumer surplus, which is not (necessarily) the case here, as can be seen by comparing (10) and (11).

The following implications of Theorem 1 follow immediately: (i) if alternative 1 is a normal good, then for fixed y and for all $p > p_0$, we have that $\bar{q}(p, y)\leq\bar{q}(p, y+p-p_0)$, and hence (10) and (11) imply that for a price increase from p_0 to p_1, $E(S^{\mathrm{EV}})\leq E(S^{\mathrm{CV}})$; (ii) for a per unit tax of τ, the average deadweight loss is given by

$$DWL^{\text{tax}}\big(S^{\text{EV}}\big) = \int_{p_0}^{p_0(1+\tau)} \bar{q}(p, y)\, dp - \tau p_0 \times \bar{q}\big(p_0(1+\tau), y\big),$$

$$DWL^{\text{tax}}\big(S^{\text{CV}}\big) = \int_{p_0}^{p_0(1+\tau)} \bar{q}(p, y + p - p_0)\, dp - \tau p_0 \times \bar{q}\big(p_0(1+\tau), y\big);$$

(iii) when a subsidy reduces prices from p_1 to p_0, the labelling of EV and CV reverses and one gets that

$$\mu^{\text{CV}}(p_0, p_1, y) = \int_{p_0}^{p_1} \bar{q}(p, y)\, dp, \tag{12}$$

$$\mu^{\text{EV}}(p_0, p_1, y) = \int_{p_0}^{p_1} \bar{q}(p, y + p - p_0)\, dp.$$

2.2. *Multinomial Choice*

In this subsection, we show that welfare analysis for unordered multinomial choice can be conducted, under an appropriate assumption (see Assumption 2 below) by reducing the problem to a binary choice problem with a composite outside option and then applying Theorem 1. Toward that end, assume that a consumer with income Y and taste η faces a mutually exclusive set of alternatives with alternative-specific prices, the classic example being the choice of mode of transportation (e.g., bus, train, walk, etc.). The consumer can pick only one among the various alternatives. Let the set of alternatives be denoted by $\{0, 1, \dots, J\}$ with P_j denoting the price of the jth alternative for $j = 1, \dots, J$ and the 0th alternative denoting not choosing any of the J alternatives. As before, assume that utility from choosing alternative 0 is $U_0(y, \eta)$ and choosing alternative j produces utility $U_j(Y - P_j, \eta)$. We are interested in calculating the marginal distribution of EV and CV resulting from a price increase for alternative 1 from $P_1 = p_{10}$ to $P_1 = p_{11}$, with the prices of the other alternatives held fixed at $(p_2, p_3, \dots, p_J)$ and income fixed at $Y = y$.

Now define

$$p_{-1} \stackrel{\text{def}}{=} (p_2, p_3, \dots, p_J), \tag{13}$$

$$U^*(p_{-1}, y, \eta) \stackrel{\text{def}}{=} \max\big\{U_0(y, \eta), U_2(y - p_2, \eta), \dots, U_J(y - p_J, \eta)\big\}.$$

A consumer of type η and income y, and facing prices (p_1, p_{-1}) chooses alternative 1 if and only if $U_1(y - p_1, \eta) > U^*(p_{-1}, y, \eta)$.

Define $\bar{q}_1(t, p_{-1}, y)$ to be the structural probability of choosing alternative 1 when its own price is t, prices of the other alternatives are p_{-1}, and income is y, that is,

$$\bar{q}_1(t, p_{-1}, y) \stackrel{\text{def}}{=} \int 1\{U_1(y-t, \eta) \geq U^*(p_{-1}, y, \eta)\}\, dF(\eta). \tag{14}$$

Define EV to be the solution S to the equation

$$\begin{aligned} &\max\{U_0(y-S, \eta), U_1(y-S-p_{10}, \eta), \\ &\qquad U_2(y-S-p_2, \eta), \ldots, U_J(y-S-p_J, \eta)\} \\ &\quad = \max\{U_0(y, \eta), U_1(y-p_{11}, \eta), \\ &\qquad\quad U_2(y-p_2, \eta), \ldots, U_J(y-p_J, \eta)\}. \end{aligned} \tag{15}$$

Similarly, define CV to be the solution S to the equation

$$\begin{aligned} &\max\{U_0(y+S, \eta), U_1(y+S-p_{11}, \eta), \\ &\qquad U_2(y+S-p_2, \eta), \ldots, U_J(y+S-p_J, \eta)\} \\ &\quad = \max\{U_0(y, \eta), U_1(y-p_{10}, \eta), \\ &\qquad\quad U_2(y-p_2, \eta), \ldots, U_J(y-p_J, \eta)\}. \end{aligned} \tag{16}$$

Note that using (13) and using the fact that $\max\{a, b, c\} = \max\{a, \max\{b, c\}\}$, (15) and (16) can be rewritten, respectively, as

$$\begin{aligned} &\max\{U^*(p_{-1}, y-S, \eta), U_1(y-S-p_{10}, \eta)\} \\ &\quad = \max\{U^*(p_{-1}, y, \eta), U_1(y-p_{11}, \eta)\}, \end{aligned} \tag{17}$$

$$\begin{aligned} &\max\{U^*(p_{-1}, y+S, \eta), U_1(y+S-p_{11}, \eta)\} \\ &\quad = \max\{U^*(p_{-1}, y, \eta), U_1(y-p_{10}, \eta)\}, \end{aligned} \tag{18}$$

which are basically (5) and (6) with $U_0(\cdot, \eta)$ replaced by the utility of the composite alternative $U^*(p_{-1}, \cdot, \eta)$. Therefore, we can perform welfare analysis in this case essentially by applying Theorem 1, as long as $U^*(p_{-1}, \cdot, \eta)$ satisfies the same assumption as $U_0(\cdot, \eta)$ in Assumption 1. Toward that end, make the following assumption.

ASSUMPTION 2: *The utility $U_j(\cdot, \eta)$ is continuous and strictly increasing for each η, for $j = 0, 1, \ldots, J$.*

Then we have the following result.

THEOREM 2: *Suppose Assumption 2 holds. Consider a price increase for alternative* 1 *from* p_{10} *to* p_{11}, *with the prices of all other alternatives held fixed at* p_{-1}. *Then EV and CV evaluated at hypothetical income* y *have marginal distributions given by*

$$\Pr[S^{\mathrm{EV}}(y, p_{10}, p_{11}, p_{-1}, \eta) \le r] = \begin{cases} 0 & \text{if } r < 0, \\ 1 - \bar{q}_1(p_{10} + r, p_{-1}, y) & \text{if } 0 \le r < p_{11} - p_{10}, \\ 1 & \text{if } r \ge p_{11} - p_{10}, \end{cases} \tag{19}$$

$$\Pr[S^{\mathrm{CV}}(y, p_{10}, p_{11}, p_{-1}, \eta) \le r] = \begin{cases} 0 & \text{if } r < 0, \\ 1 - \bar{q}_1(p_{10} + r, p_{-1}, y + r) & \text{if } 0 \le r < p_{11} - p_{10}, \\ 1 & \text{if } r \ge p_{11} - p_{10}, \end{cases} \tag{20}$$

with $\bar{q}_1(\cdot, \cdot, \cdot)$ *defined in* (14).

Analogous to Corollary 1, it follows that the expected values of EV and CV evaluated at income y are now given, respectively, by

$$E(EV) = \int_{p_{10}}^{p_{11}} \bar{q}_1(r, p_{-1}, y)\, dr, \tag{21}$$

$$E(CV) = \int_{p_{10}}^{p_{11}} \bar{q}_1(r, p_{-1}, y + r - p_{10})\, dr. \tag{22}$$

Computational Issues

As in any nonparametric estimation problem, we can estimate $\bar{q}_1(t, p_{-1}, y)$ without any parametric assumptions for those price–income combinations that lie within the range of the observed data. For combinations that may not be exactly observed in the sample, but nonetheless lie within the observed range of values, one would typically use a local smoothing method such as Nadaraya–Watson regression. However, if the hypothetical initial and/or final price lies outside the range of observed prices or if P_1 does not vary enough across individuals for given P_{-1}, Y, then either we need to use a parametric model or we can only bound the expected EV/CV (e.g., using that $\bar{q}_1(\cdot, p_{-1}, y)$ is nonincreasing). Note, however, that *these issues do not affect the conclusions of Theorems 1 or 2*—for example, expected EV is still $\int_{p_{10}}^{p_{11}} \bar{q}_1(p, p_{-1}, y)\, dp$; it is the calculation of $\bar{q}_1(p_1, p_{-1}, y)$ for (p_1, p_{-1}, y) lying outside observed price–income ranges or P_1 having limited variation that cannot be done entirely nonparametrically.

Even when one is willing to make parametric assumptions, results like (19) and (20) are still very useful, since one can apply them to compute welfare distributions directly from choice probabilities without computing expenditure functions. For example, consider the case of $(J+1)$ alternatives $\{0, 1, 2, \ldots, J\}$, linear utilities $U_j(a, \eta) = \beta_j a + \varepsilon_j$, with ε's independent and identically distributed (IID) and extreme valued, independently of price and income. Then the conditional choice probability for alternative 1 is given by the multinomial logit form[6]

$$\bar{q}_1(p_1, p_{-1}, y) = \exp\bigl(\beta_1(y - p_1)\bigr) \Big/ \Bigl(\exp\bigl(\beta_1(y - p_1)\bigr) + \sum_{j \neq 1} \exp\bigl(\beta_j(y - p_j)\bigr)\Bigr), \tag{23}$$

where $p_0 \equiv 0$ and the β's are estimable via maximum likelihood (ML). Accordingly, from (22),

$$E(S^{\mathrm{CV}}) = \int_{p_{10}}^{p_{11}} \Bigl(\exp\bigl(\beta_1(y - p_{10})\bigr) \Big/ \Bigl(\exp\bigl(\beta_1(y - p_{10})\bigr) + \sum_{j \neq 1} \exp\bigl(\beta_j(y + r - p_{10} - p_j)\bigr)\Bigr)\Bigr)\, dr. \tag{24}$$

These integrals can be calculated numerically by averaging the estimated integrand over a grid, $r_k = p_{10} + \frac{k}{K+1}(p_{11} - p_{10})$, $k = 0, \ldots, K$, that is,

$$\hat{E}(S^{\mathrm{CV}}) = \frac{1}{K+1} \sum_{k=0}^{K} \Bigl(\exp\bigl(\hat{\beta}_1(y - p_{10})\bigr) \Big/ \Bigl(\exp\bigl(\hat{\beta}_1(y - p_{10})\bigr) + \sum_{j \neq 1} \exp\Bigl(\hat{\beta}_j\Bigl(y + \frac{k}{K+1}(p_{11} - p_{10}) - p_j\Bigr)\Bigr)\Bigr)\Bigr),$$

where $\{\hat{\beta}_j\}$, $j = 0(1)J$, denotes maximum likelihood estimators (MLEs) of the β_j's in the model given by (23). Analogous computations can be performed for

[6]As pointed out by a referee, when income is endogenous, the choice probabilities conditional on a control function may no longer be expressible in the convenient logit form with a simple adjustment term. In Bhattacharya and Lee (2014), we explore estimation and inference under endogeneity in greater detail.

other distributional specifications, such as multivariate normal or McFadden and Train's (2000) mixed logit. In an actual application, one would be well advised to report numerical results for both parametric specifications as well as nonparametric ones, whenever possible, and to examine sensitivity of the findings to the extent of smoothing.

Inference

The present paper is mainly concerned with identification; a full-scale treatment of inference covering both the exogenous and endogenous income cases is being pursued in a separate paper (Bhattacharya and Lee (2014)). We simply note here that once a nonparametric estimate of $\bar{q}_1(p_1, p_{-1}, y)$ is obtained, say, using kernels or series methods, one can estimate the distribution functions of EV and CV by direct plug-in to (19) and (20). For inference on expected welfare (21) and (22), obtained by integrating the $\bar{q}_1(p_1, p_{-1}, y)$ functions over p_1 holding y fixed, one can use distributional results for partial means (see Newey (1994), Lee (2014)). For a parametric specification of choice probabilities, the estimate of expected welfare, as is apparent from (24), is a smooth function of the MLE's of model parameters. Accordingly, one can employ the bootstrap to conduct inference on mean welfare, which is justified via the following delta method.

REMARK 2: It is interesting to note that when income is endogenous, there is a distinction between (i) the marginal distribution of welfare evaluated at hypothetical income y (the parameter of interest in the present paper) and (ii) the conditional distribution of welfare evaluated at hypothetical income y for the subpopulation whose value of realized income is y'. For EV, these are given, respectively, by

$$\pi(a, y) \stackrel{\text{def}}{=} \int 1\{S^{\text{EV}}(p_{10}, p_{11}, p_{-1}, y, \eta) \leq a\}\, dF(\eta),$$

$$\pi^c(a, y, y') \stackrel{\text{def}}{=} \int 1\{S^{\text{EV}}(p_{10}, p_{11}, p_{-1}, y, \eta) \leq a\}\, dF_{\eta|Y}(\eta|y'),$$

where $F(\cdot)$ denotes the marginal distribution of η and $F_{\eta|Y}(\cdot|y')$ denotes the conditional distribution of η for the subpopulation whose current income is y'. In a treatment effect context, $\pi(a, y)$ is analogous to the average treatment effect and $\pi^c(a, y, y')$ is analogous to the average effect of treatment on the treated. In this paper, we focus on the former parameter. Of course, if income is exogenous conditional on observed covariates (as assumed in DM75, HK99, and DK05, and in random coefficient models), then the two parameters coincide.

Sequential Choice

The above results pertain to a one-time choice and are based on data from a single cross section, which, to our knowledge, is also the setup in all existing work on empirical welfare analysis. However, as noted by a referee, many discrete decisions in reality involve sequential choices over time. Welfare analysis for such dynamic discrete choice would depend on the specifics of the setting, for example, whether future prices and income are known or uncertain, whether price change is temporary or permanent, and whether income compensation is one time or recurrent. As such, a comprehensive treatment of all such cases requires a separate paper and is left to future work. Nonetheless, we note that Theorem 1 can be used for welfare analysis in some simple situations involving sequential discrete choice, for example, when an individual chooses one among several alternatives in the initial period by paying a one-time price and evaluates the overall utility of choosing each alternative by the discounted sum of expected future returns resulting from that choice, inclusive of discrete decisions to be taken optimally in the future. In this setting, Theorem 1 could be used for welfare analysis corresponding to change in an alternative-specific price in the initial period.

3. ORDERED CHOICE

The multinomial choice scenario of the previous section can be contrasted with a situation of ordered choice, that is, where a good can be bought in discrete units of $0, 1, 2, \ldots$ *and the per unit price is the same*, no matter how many units are bought, so that a change in the unit price changes the price of all nonzero alternatives simultaneously. From the identification point of view, uniform unit price restricts the number of choice sets on which we can observe the consumers' behavior relative to the multinomial case where prices of different alternatives can vary independently of each other. Continuous choice, considered in Hausman and Newey (2014), can be viewed as a limiting case of ordered choice with uniform unit price. We will demonstrate that for ordered choice, the distributions of welfare changes are *not* point-identified in general. However, if the per unit price is allowed to be *different* depending on how many units are bought (e.g., a discount is provided for larger purchases), then it becomes possible to change, say, the unit price of buying one unit while holding the unit price of other alternatives (e.g., buying two units, three units, etc.) fixed. This case will then reduce to the multinomial case above and one would end up with a point-identification result. To take a simple example, consider two scenarios. In scenario A, one chooses between one banana, two bananas, and no banana, where the price per banana is p. In this case, the distributions of CV and EV that correspond to changing the price p will not be point-identified. In scenario B, one still chooses between one banana, two bananas, and no banana, but the price is p_1 per banana if one buys a single banana and it is p_2 per banana if one buys two bananas. In this second scenario, the distributions of CV and EV arising from changing p_1 while holding p_2 fixed (or vice versa) are point-identified since one may simply view this as a multinomial choice problem with three alternatives, where the "price" of alternative 1 (buy no banana) is zero, that of alternative 2 (buy one banana) is p_1, and that of alternative 3 (buy two bananas) is $2p_2$.

Scenario A

To see the failure of point-identification in scenario A explicitly, let $U_0(y, \eta)$, $U_1(y-p, \eta)$, and $U_2(y-2p, \eta)$ denote the utility from buying zero, one, or two bananas, respectively, where p denotes the (uniform) unit price per banana and the household's income is y. Assume that for each $j = \{0, 1, 2\}$, $U_j(\cdot, \eta)$ is strictly increasing with probability 1. From the conditional choice probabilities of alternatives 0, 1, and 2 on observed budget sets (p^*, y^*), we can identify the probabilities of the sets

$$
\begin{aligned}
(25)\qquad & A_0(p^*, y^*) \\
&\quad = \{\eta : U_0(y^*, \eta) \geq \max\{U_1(y^* - p^*, \eta), U_2(y^* - 2p^*, \eta)\}\}, \\
& A_1(p^*, y^*) \\
&\quad = \{\eta : U_1(y^* - p^*, \eta) \geq \max\{U_0(y^*, \eta), U_2(y^* - 2p^*, \eta)\}\}, \\
& A_2(p^*, y^*) \\
&\quad = \{\eta : U_2(y^* - 2p^*, \eta) \geq \max\{U_0(y^*, \eta), U_1(y^* - p^*, \eta)\}\}
\end{aligned}
$$

for different values of (p^*, y^*). It is important to note in each line of the previous display that the utilities for the three alternatives being compared are evaluated at the same (p^*, y^*). The choice probabilities cannot point-identify, for example, the probability of a set like

$$\{\eta : U_1(y^* - p^*, \eta) \geq \max\{U_0(y^*, \eta), U_2(y^* - 2p, \eta)\}\},$$

where $p \neq p^*$, because no individual in the data faces two different unit prices for alternative 1 and alternative 2 in scenario A. In other words, no consumer in the population may be observed to make a choice among the three bundles $(0, y^*)$, $(1, y^* - p^*)$, and $(2, y - 2p)$ if $p \neq p^*$.

Now, suppose the per unit price p changes from p_0 to p_1, where $p_1 > p_0$. Then the resulting CV is the solution S^{CV} to the equation

$$
\begin{aligned}
(26)\qquad & \max\{U_0(y, \eta), U_1(y - p_0, \eta), U_2(y - 2p_0, \eta)\} \\
&\quad = \max\{U_0(y + S^{\mathrm{CV}}, \eta), U_1(y + S^{\mathrm{CV}} - p_1, \eta), \\
&\qquad\qquad U_2(y + S^{\mathrm{CV}} - 2p_1, \eta)\}.
\end{aligned}
$$

Consider the probability that $S^{\mathrm{CV}} = p_1 - p_0$. If the utility differences are continuously distributed, then the only situation where $S^{\mathrm{CV}} = p_1 - p_0$ is where the maximum on the LHS of (26) is $U_1(y - p_0, \eta)$ and that on the RHS is $U_1(y + S^{\mathrm{CV}} - p_1, \eta)$.[7] Thus

$$
\begin{aligned}
(27)\qquad & \Pr\bigl[S^{\mathrm{CV}} = p_1 - p_0\bigr] \\
&= \Pr\Bigl[U_1(y - p_0, \eta) \geq \max\bigl\{U_0(y, \eta), U_2(y - 2p_0, \eta)\bigr\}, \\
&\qquad U_1(y + p_1 - p_0 - p_1, \eta) \geq \max\bigl\{U_0(y + p_1 - p_0, \eta), \\
&\qquad\qquad U_2(y + p_1 - p_0 - 2p_1, \eta)\bigr\}\Bigr] \\
&= \Pr\Bigl[U_1(y - p_0, \eta) \geq \max\bigl\{U_0(y, \eta), U_2(y - 2p_0, \eta)\bigr\}, \\
&\qquad U_1(y - p_0, \eta) \geq \max\bigl\{U_0(y + p_1 - p_0, \eta), \\
&\qquad\qquad U_2(y - p_0 - p_1, \eta)\bigr\}\Bigr] \\
&= \Pr\bigl[U_1(y - p_0, \eta) \geq \max\bigl\{U_0(y + p_1 - p_0, \eta), U_2(y - 2p_0, \eta)\bigr\}\bigr]
\end{aligned}
$$

by strict monotonicity of $U_0(\cdot, \eta)$ and $U_2(\cdot, \eta)$. For standard parametric models (e.g., $U_j(y, \eta) = \alpha_j \ln(y) + \eta_j$ with η_j scalar and normally distributed), the probability in (27) is positive. However, it is clear from (27) that the probability that $S^{\mathrm{CV}} = p_1 - p_0$ equals the probability of choosing the bundle $(1, y - p_0)$ over the bundles $(0, y + p_1 - p_0)$ and $(2, y - 2p_0)$. The latter probability can be nonparametrically point-identified if and only if some consumers in the population face the choice among these three bundles, which can happen if and only if for some (p^*, y^*), we have that

$$
\begin{aligned}
& y^* - p^* = y - p_0, \\
& y^* = y + p_1 - p_0, \\
& y^* - 2p^* = y - 2p_0.
\end{aligned}
$$

Replacing the second equation in the first yields $p^* = y + p_1 - p_0 - y + p_0 = p_1$ and replacing this in the first equation yields $y^* = y + p_1 - p_0$. But $y^* = y + p_1 - p_0$ and $p^* = p_1$ does not satisfy the third equation. Thus, there is

[7]For example, if instead the first term is the maximum for the LHS of (26) and the second term is the maximum for the RHS of (26) with $S = p_1 - p_0$, then

$$
U_0(y, \eta) = U_1\bigl(y + \overbrace{p_1 - p_0}^{S} - p_1, \eta\bigr) = U_1(y - p_0, \eta),
$$

which will have zero probability if $U_1(y - p_0, \eta) - U_0(y, \eta)$ is continuously distributed, for example, for $j = 0, 1, 2$, $U_j(y - p_j, \eta) = \alpha_j \ln(y - p_j) + \eta_j$ with $\{\eta_0, \eta_1, \eta_2\}$ having extreme value or joint normal distribution.

no (p^*, y^*) that satisfies all three equations simultaneously. In other words, there cannot be any consumer in the population who may be observed to make a choice among the bundles $(0, y + p_1 - p_0)$, $(1, y - p_0)$, and $(2, y - 2p_0)$, so that we cannot identify the probability of choosing the bundle $(1, y - p_0)$ over the bundles $(0, y + p_1 - p_0)$ and $(2, y - 2p_0)$, which is the probability of $S^{\mathrm{CV}} = p_1 - p_0$. Thus, the probability that $S^{\mathrm{CV}} = p_1 - p_0$ is positive, but it is nonparametrically *un*identified, so the distribution of the CV cannot be nonparametrically point-identified.

Scenario B

Now, consider scenario B. Indeed, if the *per unit* price when two units are consumed is fixed at p_2 and the unit price for consuming one unit rises from p_0 to p_1, then

$$\begin{aligned}
&\Pr\big[S^{\mathrm{CV}} = p_1 - p_0\big] \\
&\quad = \Pr\Big[U_1(y - p_0, \eta) \geq \max\big\{U_0(y, \eta), U_2(y - 2p_2, \eta)\big\}, \\
&\qquad\qquad U_1(y + p_1 - p_0 - p_1, \eta) \geq \max\big\{U_0(y + p_1 - p_0, \eta), \\
&\qquad\qquad\qquad\qquad U_2(y + p_1 - p_0 - 2p_2, \eta)\big\}\Big] \\
&\quad = \Pr\Big[U_1(y - p_0, \eta) \geq \max\big\{U_0(y, \eta), U_2(y - 2p_2, \eta)\big\}, \\
&\qquad\qquad U_1(y - p_0, \eta) \geq \max\big\{U_0(y + p_1 - p_0, \eta), \\
&\qquad\qquad\qquad\qquad U_2(y + p_1 - p_0 - 2p_2, \eta)\big\}\Big] \\
&\quad = \Pr\Big[U_1(y - p_0, \eta) \geq \max\big\{U_0(y + p_1 - p_0, \eta), \\
&\qquad\qquad\qquad\qquad U_2(y + p_1 - p_0 - 2p_2, \eta)\big\}\Big] \\
&\quad = \Pr\Big[U_1\big((y + p_1 - p_0) - p_1, \eta\big) \geq \max\big\{U_0(y + p_1 - p_0, \eta), \\
&\qquad\qquad\qquad\qquad U_2(y + p_1 - p_0 - 2p_2, \eta)\big\}\Big] \\
&\quad = \bar{q}_1(p_1, p_2, y + p_1 - p_0),
\end{aligned}$$

which is point-identified from the observed choice probability of alternative 1 at price p_1, income $y + p_1 - p_0$, and unit price of alternative 2 (i.e., of buying 2 units) fixed at p_2.

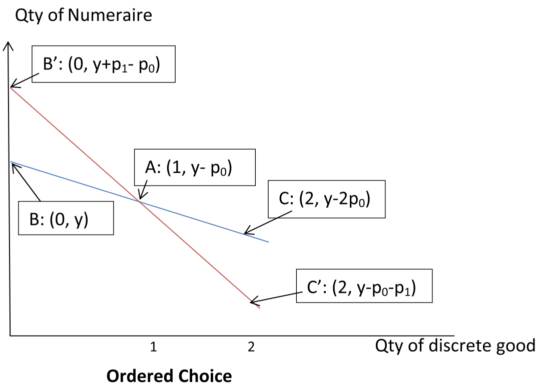

FIGURE 1.—CV for ordered choice.

Graphical Illustration

The above discussion is graphically depicted in the following figures. First, considered ordered choice, depicted in Figure 1.

At price p_0 and income y, individuals choose between the bundles $B = (0, y)$, $A = (1, y - p_0)$, and $C = (2, y - 2p_0)$; at price p_1 and compensated income $y + p_1 - p_0$, they choose between the bundles $B' = (0, y + p_1 - p_0)$, $A = (1, y + p_1 - p_0 - p_1) = (1, y - p_0)$, and $C' = (2, y + p_1 - p_0 - 2p_1) = (2, y - p_1 - p_0)$. By monotonicity of utility, for every η, $B' \succ B$ and $C \succ C'$. Therefore,

$$\begin{aligned}\Pr\bigl(S^{\mathrm{CV}} = p_1 - p_0\bigr) &= \Pr\bigl[\{A \succ B, A \succ C\} \cap \bigl\{A \succ B', A \succ C'\bigr\}\bigr] \\ &= \Pr\bigl(A \succ C, A \succ B'\bigr).\end{aligned}$$

But in the ordered choice case, we observe choice behavior only across bundles that lie on the same straight line (e.g., (B, A, C) or (B', A, C')). Since (B', A, C) do not lie on a straight line, we cannot point-identify the probability $\Pr(A \succ C, A \succ B')$. Now consider unordered choice, Figure 2.

With the price of alternative 2 fixed at p_2, we observe choice behavior across bundles of the form

$$C(z, p, p_2) := \bigl\{(0, z), (1, z - p), (2, z - 2p_2)\bigr\}$$

for every combination of (z, p). Now

$$\begin{aligned}\Pr\bigl(S^{\mathrm{CV}} = p_1 - p_0\bigr) &= \Pr\bigl[\{A \succ B, A \succ D\} \cap \bigl\{A \succ B', A \succ D'\bigr\}\bigr] \\ &= \Pr\bigl(A \succ D', A \succ B'\bigr),\end{aligned} \tag{28}$$

since $D' \succ D$ by monotonicity. Therefore, we can compute the probability (28) as the probability of choosing $(1, z - p)$ from $C(z, p, p_2)$, with $z = y + p_1 - p_0$ and $p = p_1$.

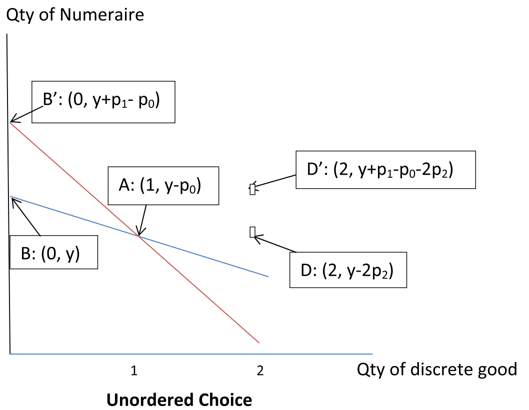

FIGURE 2.—CV for unordered choice.

REMARK 3: It is interesting to note that if the functional forms of the utility functions were parametrically specified, and the dimension and distribution of heterogeneity were assumed to be known (as in HK99 or DK05), then the probability in (27) would appear to be point-identified. For example, if one assumes that for $j = 0, 1, 2$, $U_j(y, \eta) = \alpha_j \ln(y) + \eta_j$ with $\{\eta_0, \eta_1, \eta_2\}$ distributed as IID and extreme valued, independent of price and income, then (27) reduces to

$$\begin{aligned}
&\Pr\Big[U_1(y - p_0, \eta) \geq \max\big\{U_0(y + p_1 - p_0, \eta), U_2(y - 2p_0, \eta)\big\}\Big] \\
&\quad = \Pr\Big[\alpha_1 \ln(y - p_0) + \eta_1 \geq \max\big\{\alpha_0 \ln(y + p_1 - p_0) + \eta_0, \\
&\qquad\qquad\qquad\qquad\qquad\qquad \alpha_2 \ln(y - 2p_0) + \eta_2\big\}\Big] \\
&\quad = \exp\big(\alpha_1 \ln(y - p_0)\big) \\
&\qquad \big/\big(\exp\big(\alpha_0 \ln(y + p_1 - p_0)\big) + \exp\big(\alpha_1 \ln(y - p_0)\big) \\
&\qquad\quad + \exp\big(\alpha_2 \ln(y - 2p_0)\big)\big),
\end{aligned}$$

which can be readily "calculated" using ML estimates of the α's obtained from choice data. However, the identification result underlying this calculation is artificial in that it is driven entirely by functional form assumptions. In other words, nonparametric point-identifiability of welfare distributions in the case of unordered choice and its failure in the ordered case is a fundamental difference that would not be apparent if one only considered parametric models.

4. SUMMARY AND CONCLUSION

The key insight of this paper is that for price change in a binary or unordered multinomial choice situation, the choice probabilities alone contain all the relevant information for nonparametric recovery of the resulting welfare distributions. This result is valid under unrestricted forms of unobserved heterogeneity and utility functions, and continues to hold even if the dimension—and, therefore, the distribution—of unobserved heterogeneity is neither specified nor identified. Interestingly, the result fails in the case of ordered choice with three or more alternatives if the unit price is required to be the same no matter how many units are bought, thereby providing a link with Hausman and Newey's (2014) recent finding that for price change of a *continuous* good, averages of money-metric welfare are only set-identified under unrestricted heterogeneity.

On the practical end, the distributions of welfare are expressed here as simple closed-form transformations of choice probabilities, enabling easy computation and inference. Even if one approximates the choice probabilities by a parametric model, for example, a mixed logit, these closed-form expressions can be used to compute welfare distributions directly from these choice probabilities without requiring one to compute expenditure functions by reverting to potentially misspecified utility functions and heterogeneity distributions.

To our knowledge, the present paper delivers the first set of results in the econometric literature on nonparametric welfare analysis for discrete choice and does so for essentially unrestricted preference distributions. Consequently, it significantly advances the existing literature that either (a) ignored the key identification problem by assuming negligible income effects or (b) based welfare calculations on parametrically specified—and, consequently, potentially misspecified—utility functions and heterogeneity distributions without recognizing that the conditional choice probabilities themselves contain all the relevant information for money-metric welfare analysis under essentially unrestricted preference distributions.

APPENDIX

PROOF OF LEMMA 1: We denote the individual EV $S^{\mathrm{EV}}(y, p_0, p_1, \eta)$ simply by S to avoid cumbersome notation in the proof. Also for ease of reference, we rewrite (5) again:

$$\begin{aligned} &\max\{U_0(y-S,\eta), U_1(y-S-p_0,\eta)\} \\ &\quad = \max\{U_0(y,\eta), U_1(y-p_1,\eta)\}. \end{aligned}$$

By monotonicity of $U_1(\cdot,\eta)$ and $U_0(\cdot,\eta)$, we must have that $S \geq 0$ in order for (5) to hold.

Now, in case (i), $0 \le U_0(y,\eta) - U_1(y-p_0,\eta)$. So if $S > 0$, then

$$\begin{aligned}&\max\{U_0(y-S,\eta), U_1(y-S-p_0,\eta)\}\\&\quad< \max\{U_0(y,\eta), U_1(y-p_0,\eta)\}\\&\quad= U_0(y,\eta)\\&\quad\le \max\{U_0(y,\eta), U_1(y-p_1,\eta)\},\end{aligned}$$

contradicting (5). This implies that $S = 0$.

Now, consider case (ii). Given the restriction $0 \le U_0(y,\eta) - U_1(y-p_1,\eta)$, the RHS of (5) is $U_0(y,\eta)$. Therefore, from (5), S must satisfy

$$U_0(y,\eta) = \max\{U_1(y-S-p_0,\eta), U_0(y-S,\eta)\}. \tag{29}$$

Now, (29) is equivalent to

$$U_0(y,\eta) = U_1(y-S-p_0,\eta). \tag{30}$$

To see why, suppose the first term on the RHS of (29) is smaller than the second, that is,

$$U_1(y-S-p_0,\eta) < U_0(y-S,\eta). \tag{31}$$

Then (29) implies $U_0(y,\eta) = U_0(y-S,\eta)$, implying $S = 0$ by strict monotonicity of $U_0(\cdot,\eta)$. But then the first term on the RHS of (29), namely, $U_1(y-S-p_0,\eta)$, equals $U_1(y-p_0,\eta)$, while the second term equals $U_0(y,\eta)$; but because we know that $U_1(y-p_0,\eta) \ge U_0(y,\eta)$ (we are in case (ii) of the proposition), the inequality (31) is violated. Therefore, we must have that $U_1(y-S-p_0,\eta) \ge U_0(y-S,\eta)$, so that the maximum on the RHS of (29) must equal $U_1(y-S-p_0,\eta)$, whence the conclusion (30) follows.

From (30), using monotonicity of $U_1(\cdot,\eta)$, we have that $S = y - p_0 - U_1^{-1}(U_0(y,\eta),\eta)$. Note that by the continuity condition of Assumption 1, the inverse $U_1^{-1}(\cdot,\eta)$ is defined everywhere.

Finally, consider case (iii): Given the restriction $0 > U_0(y,\eta) - U_1(y-p_1,\eta)$, the RHS of (5) is $U_1(y-p_1,\eta)$. Now, suppose the LHS of (5) is $U_0(y-S,\eta)$. But since $S \ge 0$, we must have that

$$U_0(y-S,\eta) \le U_0(y,\eta) < U_1(y-p_1,\eta) = U_0(y-S,\eta) \quad \text{by (5)},$$

a contradiction. Therefore, the LHS of (5) must be $U_1(y-S-p_0,\eta)$ and, therefore, by (5), $U_1(y-S-p_0,\eta) = U_1(y-p_1,\eta)$, whence, by strict monotonicity of $U_1(\cdot,\eta)$, we get that $S = p_1 - p_0$. *Q.E.D.*

PROOF OF LEMMA 2: We denote the individual CV $S^{\mathrm{CV}}(y,p_0,p_1,\eta)$ simply by S to avoid cumbersome notation in the proof. Also for ease of reference, we rewrite (6) again:

$$\max\{U_0(y+S,\eta), U_1(y+S-p_1,\eta)\}$$
$$= \max\{U_0(y,\eta), U_1(y-p_0,\eta)\}.$$

First observe that by (6), we must have that $S \geq 0$. Otherwise, the LHS of (6) must be strictly smaller than the RHS, by the monotonicity of $U_0(\cdot)$ and $U_1(\cdot)$. Now consider the following cases.

Case (i): We have $0 \leq U_0(y,\eta) - U_1(y-p_0,\eta)$.

Since $0 \leq U_0(y,\eta) - U_1(y-p_0,\eta)$, then the RHS of (6) is $U_0(y,\eta)$. If $S > 0$, then the first term on the LHS of (6) must be strictly larger than $U_0(y,\eta)$, by strict monotonicity of $U_0(\cdot,\eta)$. This would imply that the LHS of (6) must be strictly larger than $U_0(y,\eta)$—a contradiction. Therefore, in this case, we must have $S = 0$. Intuitively, this means that those η who were not buying at the initial price p_0 do not need to be compensated.

Now, suppose case (i) does not hold, so that the RHS maximum is, in fact, $U_1(y-p_0,\eta)$, that is, $U_0(y,\eta) - U_1(y-p_0,\eta) < 0$. This corresponds to those η's who buy the good at price p_0. Now there are two possibilities regarding which term is the maximum in the LHS of (6): case (ii) corresponds to when the maximum is the first term and case (iii) corresponds to when the maximum is the second term.

Case (ii): Accordingly, first assume that the LHS maximum is $U_0(y+S,\eta)$. Then S must satisfy

$$U_0(y+S,\eta) = U_1(y-p_0,\eta) \quad \Longrightarrow \quad S = U_0^{-1}\big(U_1(y-p_0,\eta),\eta\big) - y. \tag{32}$$

Note that by the continuity condition of Assumption 1, the inverse $U_0^{-1}(\cdot,\eta)$ is defined everywhere. So that (32) simultaneously satisfies that the LHS maximum of (6) is $U_0(y+S,\eta)$, we need

$$\begin{aligned}
& U_0(y+S,\eta) \geq U_1(y+S-p_1,\eta) \\
& \overset{\text{substituting } S \text{ from (32)}}{\Longrightarrow} \quad U_1(y-p_0,\eta) \geq U_1(y+S-p_1,\eta) \\
& \Longrightarrow \quad y - p_0 \geq y + S - p_1 \\
& \Longrightarrow \quad p_1 - p_0 \geq S = U_0^{-1}\big(U_1(y-p_0,\eta),\eta\big) - y \\
& \Longrightarrow \quad U_0(y+p_1-p_0,\eta) \geq U_1(y-p_0,\eta) \\
& \Longrightarrow \quad 0 \leq U_0(y+p_1-p_0,\eta) - U_1(y-p_0,\eta).
\end{aligned}$$

Thus, we arrive at the conclusion of case (ii), namely,

$$U_0(y) - U_1(y-p_0) < 0 \leq U_0(y+p_1-p_0,\eta) - U_1(y-p_0,\eta) \quad \text{and}$$
$$S = U_0^{-1}\big(U_1(y-p_0,\eta),\eta\big) - y.$$

Case (iii): Finally, consider the remaining case where the maximum of the LHS of (6) is $U_1(y+S-p_1,\eta)$, whence we have

$$U_1(y+S-p_1,\eta)=U_1(y-p_0,\eta) \quad \Longrightarrow \quad S=p_1-p_0.$$

Replacing S in the LHS of (6), so as to be consistent with our assumption that the LHS maximum is $U_1(y+S-p_1,\eta)$, we must have that

$$\begin{aligned}&U_1(y+S-p_1,\eta)\geq U_0(y+S,\eta)\\ &\quad\Longleftrightarrow\quad U_1(y-p_0,\eta)\geq U_0(y+p_1-p_0,\eta),\end{aligned}$$

which is precisely case (iii) of the proposition, namely, $U_1(y-p_0,\eta)\geq U_0(y+p_1-p_0,\eta)$. *Q.E.D.*

PROOF OF THEOREM 1: First consider EV. The compensation must be nonnegative and no larger than p_1-p_0; otherwise (5) will be violated. EV is zero for those not purchasing at p_0 and, hence, EV has a point mass equal to the probability of no purchase at p_0, which is given by $1-\bar{q}(p_0,y)$. So the only nontrivial step is for $0<a<p_1-p_0$. This case corresponds to case (ii) of Lemma 1. Accordingly, for $0<a<p_1-p_0$, the probability of compensation not exceeding a is given by

$$\begin{aligned}&\Pr\bigl(S^{\mathrm{EV}}=0,0\leq U_0(y,\eta)-U_1(y-p_0,\eta)\bigr)\\ &\qquad+\Pr\bigl(y-p_0-U_1^{-1}\bigl(U_0(y,\eta),\eta\bigr)\leq a,\\ &\qquad\qquad U_0(y,\eta)-U_1(y-p_0,\eta)<0\leq U_0(y,\eta)-U_1(y-p_1,\eta)\bigr)\\ &\quad=1-\bar{q}(p_0,y)\\ &\qquad+\Pr\bigl(0\leq U_0(y,\eta)-U_1(y-p_0-a,\eta),\\ &\qquad\qquad U_0(y,\eta)-U_1(y-p_0,\eta)<0\leq U_0(y,\eta)-U_1(y-p_1,\eta)\bigr)\\ &\quad=1-\bar{q}(p_0,y)\\ &\qquad+\Pr\bigl(U_0(y,\eta)-U_1(y-p_0,\eta)\\ &\qquad\qquad<0\leq U_0(y,\eta)-U_1(y-p_0-a,\eta)\bigr),\quad \text{since } a<p_1-p_0\\ &\quad=1-\bar{q}(p_0,y)+1-\bar{q}(p_0+a,y)-\bigl(1-\bar{q}(p_0,y)\bigr)\\ &\quad=1-\bar{q}(p_0+a,y),\end{aligned}$$

as desired.

Now consider CV. First recall from (3) that for any $a>0$, we have that

$$\begin{aligned}&\bar{q}(p_0,y)=\Pr\bigl(0>U_0(y,\eta)-U_1(y-p_0,\eta)\bigr),\\ &\bar{q}(a+p_0,y+a)=\Pr\bigl(0>U_0(y+a,\eta)-U_1(y-p_0,\eta)\bigr),\end{aligned}$$

where the probabilities are computed with respect to (w.r.t.) the marginal distribution of η.

Now, from Lemma 2 and its proof, it is clear that for any η, the compensation is nonnegative and it equals zero for those not buying at price p_0. Hence, the CV has no mass below 0 and a point mass of $1-\bar{q}(p_0, y)$ at 0. Also, it is clear that the compensation cannot exceed $p_1 - p_0$ for any η; otherwise, (6) will be violated. Therefore, the CDF of CV must reach 1 at $p_1 - p_0$. So the only nontrivial case is $0 < a < p_1 - p_0$. This corresponds to case (ii) of Lemma 2. Accordingly, for $0 < a < p_1 - p_0$, the probability of the compensation being no larger than a is given by

$$
\begin{aligned}
(33)\qquad & \Pr\bigl(S^{\mathrm{CV}} = 0, U_0(y,\eta) - U_1(y-p_0,\eta) \geq 0\bigr) \\
&\quad + \Pr\bigl\{U_0^{-1}\bigl(U_1(y-p_0,\eta),\eta\bigr) - y < a, \\
&\qquad U_0(y,\eta) - U_1(y-p_0,\eta) \\
&\qquad < 0 \leq U_0(y+p_1-p_0,\eta) - U_1(y-p_0,\eta)\bigr\} \\
&= 1 - \bar{q}(p_0, y) \\
&\quad + \Pr\bigl\{U_0^{-1}\bigl(U_1(y-p_0,\eta),\eta\bigr) - y < a, \\
&\qquad U_0(y,\eta) - U_1(y-p_0,\eta) \\
&\qquad < 0 \leq U_0(y+p_1-p_0,\eta) - U_1(y-p_0,\eta)\bigr\}.
\end{aligned}
$$

Now

$$
\begin{aligned}
& \Pr\bigl(U_0^{-1}\bigl(U_1(y-p_0,\eta),\eta\bigr) - y < a, \\
&\quad U_0(y,\eta) - U_1(y-p_0,\eta) \\
&\quad < 0 \leq U_0(y+p_1-p_0,\eta) - U_1(y-p_0,\eta)\bigr) \\
&= \Pr\bigl(0 < U_0(y+a,\eta) - U_1(y-p_0,\eta), \\
&\qquad U_0(y,\eta) - U_1(y-p_0,\eta) \\
&\qquad < 0 \leq U_0(y+p_1-p_0,\eta) - U_1(y-p_0,\eta)\bigr) \\
&= \Pr\bigl(U_0(y,\eta) - U_1(y-p_0,\eta) \\
&\qquad < 0 \leq U_0(y+a,\eta) - U_1(y-p_0,\eta)\bigr), \quad \text{since } a < p_1 - p_0 \\
&= \Pr\bigl(0 \leq U_0(y+a,\eta) - U_1(y-p_0,\eta)\bigr) \\
&\quad - \Pr\bigl(U_0(y,\eta) - U_1(y-p_0,\eta) \geq 0, \\
&\qquad U_0(y+a,\eta) - U_1(y-p_0,\eta) \geq 0\bigr) \\
&= \Pr\bigl(0 < U_0(y+a,\eta) - U_1(y-p_0,\eta)\bigr) \\
&\quad - \Pr\bigl(0 \leq U_0(y,\eta) - U_1(y-p_0,\eta)\bigr), \quad \text{since } a \geq 0
\end{aligned}
$$

$$
\begin{aligned}
&= \Pr\big(0 < U_0(y+a,\eta) - U_1\big(y+a-(a+p_0),\eta\big)\big) \\
&\quad - \Pr\big(0 \le U_0(y,\eta) - U_1(y-p_0,\eta)\big) \\
&= \Pr\big(0 > U_0(y,\eta) - U_1(y-p_0,\eta)\big) \\
&\quad - \Pr\big(0 > U_0(y+a,\eta) - U_1\big(y+a-(a+p_0),\eta\big)\big) \\
&= \bar{q}(p_0,y) - \bar{q}(a+p_0,y+a).
\end{aligned}
$$

Substituting in (33), we get that for $0 < a < p_1 - p_0$,

$$
\begin{aligned}
\Pr\{S^{\mathrm{CV}} \le a\} &= 1 - \bar{q}(p_0,y) + \bar{q}(p_0,y) - \bar{q}(a+p_0,y+a) \\
&= 1 - \bar{q}(a+p_0,y+a),
\end{aligned}
$$

as desired. *Q.E.D.*

PROOF OF COROLLARY 1: We use the well known result (see Karr (1993, p. 113)) that for a positive random variable X with CDF $F_X(\cdot)$, the expectation is given by

$$
E(X) = \int_0^\infty \big(1 - F_X(a)\big)\,da. \tag{34}
$$

From (7) and (34), the expected EV is given by

$$
\begin{aligned}
\mu^{\mathrm{EV}}(y,p_0,p_1) &= \int_0^{p_1-p_0} \bar{q}(a+p_0,y)\,da \\
&= \int_{p_0}^{p_1} \bar{q}(z,y)\,dz \quad \text{by change of variables } z = a+p_0.
\end{aligned}
$$

Next, using (8) and (34), we get that the expected CV is given by

$$
\begin{aligned}
\mu^{\mathrm{CV}}(y,p_0,p_1) &= \int_0^{p_1-p_0} \bar{q}(a+p_0,y+a)\,da \\
&= \int_{p_0}^{p_1} \bar{q}(z,y+z-p_0)\,dz \quad \text{substituting } z = a+p_0.
\end{aligned}
$$

Q.E.D.

PROOF OF THEOREM 2: We provide the proof for CV; the proof of EV is exactly analogous. Recall from (18) that the compensating variation is the solution S to the equation

$$
\begin{aligned}
&\max\big\{U^*(y+S,p_{-1},\eta), U_1(y+S-p_{11},\eta)\big\} \\
&\quad = \max\big\{U^*(y,p_{-1},\eta), U_1(y-p_{10},\eta)\big\},
\end{aligned} \tag{35}
$$

where $p_{-1} = (p_2, p_3, \ldots, p_J)$ and

$$U^*(y, p_{-1}, \eta) \overset{\text{def}}{=} \max\{U_0(y, \eta), U_2(y - p_2, \eta), \ldots, U_J(y - p_J, \eta)\}.$$

Equation (35) is exactly analogous to (6) with $U_0(y, \eta)$ replaced by $U^*(y, p_{-1}, \eta)$.

Assumption 2 implies that $U^*(\cdot, p_{-1}, \eta)$ is strictly increasing and continuous for each η. Let $U^{*-1}(p_{-1}, b, \eta)$ denote the unique solution in x to the equation $U^*(p_{-1}, x, \eta) = b$. Note that by the continuity condition of Assumption 2, the inverse $U^{*-1}(p_{-1}, \cdot, \eta)$ is defined everywhere. Now it can be seen that all our results from the binary case carry over with the utility $U_0(y, \eta)$ replaced by $U^*(y, p_{-1}, \eta)$, since the latter does not involve the price of alternative 1. Following the proof of Lemma 2 with $U_0(y, \eta)$ replaced by $U^*(y, p_{-1}, \eta)$ yields

$$S^{\mathrm{CV}} = \begin{cases} 0 & \text{if } U_1(y - p_{10}, \eta) < U^*(y, p_{-1}, \eta), \\ U^{*-1}\big(U_1(y - p_{10}, \eta), p_{-1}, \eta\big) - y & \\ & \text{if } \big(U^*(y, p_{-1}, \eta) \le U_1(y - p_{10}, \eta) \\ & \quad < U_0(y + p_{11} - p_{10}, \eta)\big), \\ p_{11} - p_{10} & \text{if } U^*(y + p_{11} - p_{10}, p_{-1}, \eta) < U_1(y - p_{10}, \eta). \end{cases}$$

Similarly, following the proof of Theorem 1 with $\bar{q}(p, y)$ replaced by $\bar{q}_1(p_1, p_{-1}, y)$ (defined in (14)) yields, analogous to (8), that

$$\begin{aligned} &\Pr\big[S^{\mathrm{CV}}(y, p_{10}, p_{11}, p_{-1}, \eta) \le r\big] \\ &\quad = \begin{cases} 0 & \text{if } r < 0, \\ 1 - \bar{q}_1(p_{10} + r, p_{-1}, y + r) & \text{if } 0 \le r < p_{11} - p_{10}, \\ 1 & \text{if } r \ge p_{11} - p_{10}. \end{cases} \end{aligned}$$

Q.E.D.

Binary Choice Example Where the Heterogeneity Dimension Is not Identified

Suppose $\eta \equiv (\eta_1, \eta_0)$ is jointly independent of price and income (P, Y) and $\eta_1 \perp \eta_0$. Assume that the support of price distribution in the data is contained in $[0, p_H]$ and income is bounded below by y_L with $y_L > p_H > 0$. Let

$$U_1(Y - P, \eta) = Y - P + \eta_1, \quad U_0(Y, \eta) = (1 - \eta_0)Y,$$

where η_0 is distributed uniform $(0, 1)$ and the support of η_1—denoted by T—is contained in $(p_H - y_L, 0)$. Denote the CDF of η_1 by $G(\cdot)$. An individual of type (y, η) and facing price p buys the good if and only if $y - p + \eta_1 > (1 - \eta_0)y$.

Thus, for any fixed $\eta = (\eta_1, \eta_0)$ in the support, the utility functions are continuous and strictly increasing in income. Thus, Theorem 1 applies and it implies that the distributions of EV and CV arising from a price change are point-identified.

Now, consider the choice probability in this model. Since $p_H - y_L \leq \eta_1 \leq 0$ with probability 1 (w.p.1), it follows that for any p, y in the support of the data, we must have that $p - y < \eta_1 < p$ or

$$0 < \frac{p - \eta_1}{y} < 1 \quad \text{w.p.1.} \tag{36}$$

Therefore, the structural choice probability of alternative 1 at price p and income y is given by

$$\begin{aligned}
\bar{q}_1(p, y) &= \Pr\{y - p + \eta_1 > (1 - \eta_2)y\} \\
&= \Pr\{\eta_2 y + \eta_1 > p\} \\
&= \Pr\left\{\eta_2 > \frac{p - \eta_1}{y}\right\}, \quad \text{since } y > 0 \\
&= \int_T \left(1 - \frac{p - \eta_1}{y}\right) dG(\eta_1) \\
&\quad \text{by } \eta_1 \perp \eta_2, \text{ inequality (36), and } \eta_2 \sim U(0,1) \\
&= \left(1 - \frac{p}{y}\right) + \frac{1}{y} E(\eta_1).
\end{aligned}$$

Thus, the choice probability $\bar{q}_1(p, y)$ depends on the distribution of η_1 only through its expectation. Therefore, all distributions for η_1 with support contained in $(p_H - y_L, 0)$ and having the same expectation will give rise to the same choice probability *for each value of* p *and* y, implying that the distribution of η_1 cannot be identified from the choice probabilities alone. In particular, $\eta_1 \sim \text{Uniform}[p_H - y_L, 0]$ and $\eta_1 = \frac{p_H - y_L}{2}$ with probability 1 will both produce identical $\bar{q}_1(p, y)$ *for all* (p, y) in the support of price and income. This implies that the dimension of heterogeneity is also not identified ($\dim(\eta_1, \eta_0)$ is 1 when η_1 is degenerate and 2 when η_1 is uniform). Yet the distributions of EV and CV are point-identified from $\bar{q}_1(\cdot, \cdot)$ as implied by Theorem 1 above.

The above example demonstrates that identifiability of the heterogeneity distribution or even correct specification of its dimension is not a requirement for identifiability of welfare distributions.

REFERENCES

BHATTACHARYA, D., AND Y. LEE (2014): "Nonparametric Inference for Welfare Distributions in Discrete Choice Models," Unpublished Manuscript. [632,633]

BLUNDELL, R., AND J. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics*. Cambridge: Cambridge University Press. [622]

BLUNDELL, R., J. HOROWITZ, AND M. PERRY (2012): "Measuring the Price Responsiveness of Gasoline Demand: Economic Shape Restrictions and Nonparametric Demand Estimation," *Quantitative Economics*, 3 (1), 29–51. [620]

BLUNDELL, R., D. KRISTENSEN, AND R. MATZKIN (2014): "Bounding Quantile Demand Functions Using Revealed Preference Inequalities," *Journal of Econometrics*, 179, 112–127. [620]

DAGSVIK, J., AND A. KARLSTROM (2005): "Compensating Variation and Hicksian Choice Probabilities in Random Utility Models That Are Nonlinear in Income," *Review of Economic Studies*, 72 (1), 57–76. [619]

DOMENCICH, T., AND D. MCFADDEN (1975): *Urban Travel Demand—A Behavioral Analysis*. Oxford: North-Holland. [618]

GAUTIER, E., AND Y. KITAMURA (2013): "Nonparametric Estimation in Random Coefficients Binary Choice Models," *Econometrica*, 81 (2), 581–607. [620,622]

HAUSMAN, J. (1981): "Exact Consumer's Surplus and Deadweight Loss," *American Economic Review*, 71 (4), 662–676. [618]

HAUSMAN, J., AND W. NEWEY (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63 (6), 1445–1476. [618]

——— (2014): "Individual Heterogeneity and Average Welfare," Working Paper 42/14, CEMMAP. [617,620,622,634,640]

HERRIGES, J., AND C. KLING (1999): "Nonlinear Income Effects in Random Utility Models," *Review of Economics and Statistics*, 81 (1), 62–72. [619]

HODERLEIN, S., AND A. VANHEMS (2011): "Welfare Analysis Using Nonseparable Models," Discussion Paper CWP01/11, CEMMAP. [620]

ICHIMURA, H., AND T. S. THOMPSON (1998): "Maximum Likelihood Estimation of a Binary Choice Model With Random Coefficients of Unknown Distribution," *Journal of Econometrics*, 86, 269–295. [620,622]

KARR, A. F. (1993): *Probability*. New York: Springer-Verlag. [645]

KITAMURA, Y., AND J. STOYE (2013): "Nonparametric Analysis of Random Utility Models: Testing," Report, Cornell and Yale University. [620]

LEE, Y. (2014): "Partial Mean Processes With Generated Regressors: Continuous Treatment Effects and Nonseparable Models," Working Paper, University of Oxford. [633]

LEWBEL, A. (2001): "Demand Systems With and Without Errors," *American Economic Review*, 91 (3), 611–618. [618]

LEWBEL, A., AND K. PENDAKUR (2013): "Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients," Report, Boston College. [620]

LEWBEL, A., O. LINTON, AND D. MCFADDEN (2011): "Estimating Features of a Distribution From Binomial Data," *Journal of Econometrics*, 162 (2), 170–188. [620]

MANSKI, C. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228. [619]

MATZKIN, R. L. (1992): "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, 60 (2), 239–270. [619]

MCFADDEN, D., AND K. RICHTER (1991): "Stochastic Rationality and Revealed Stochastic Preference," in *Preferences, Uncertainty, and Rationality*, ed. by J. Chipman, D. McFadden, and K. Richter. Summertown, Oxford: Westview Press, 161–186. [620]

MCFADDEN, D., AND K. TRAIN (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15 (5), 447–470. [633]

NEWEY, W. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253. [633]

PETRIN, A., AND K. TRAIN (2010): "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, 47 (1), 3–13. [622]

SMALL, K., AND H. ROSEN (1981): "Applied Welfare Economics With Discrete Choice Models," *Econometrica*, 49 (1), 105–130. [619]

VARTIA, Y. (1983): "Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions," *Econometrica*, 51 (1), 79–98. [618]

Dept. of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ, U.K.; debobhatta@gmail.com.

Manuscript received June, 2014; final revision received November, 2014.