Empirical Welfare Analysis for Discrete Choice: Some New Results

Debopam Bhattacharya*

University of Cambridge

September 19, 2016.

Abstract

We develop methods of empirical welfare-analysis in multinomial choice settings, under completely general consumer-heterogeneity and income-effects. Our results pertain to three practically important scenarios, viz., (i) simultaneous price-change of multiple alternatives, (ii) introduction/elimination of a choice-alternative, and (iii) choice among non-exclusive options. These do not follow from the single price-change results of Bhattacharya (Econometrica, 2015). In program-evaluation contexts, they enable estimation of *compensated* program-effects, i.e., how much the subjects themselves value a policy-intervention, and the resulting deadweight-loss. Welfare-analysis under endogeneity is briefly discussed.

Keywords: Multinomial Choice, Unobserved Heterogeneity, Nonparametric Welfare Analysis, Compensated Program Effects, Deadweight Loss, Compensating Variation, Multiple Price Change, Elimination of Alternative, Non-exclusive Choice.

1 Introduction

Welfare calculations, based on the compensating and equivalent variation (CV and EV, henceforth), lie at the heart of economic policy analysis. Although theoretically well-understood, these measures are rarely calculated or reported as part of econometric program evaluations. Indeed, empirical studies – both reduced-form and structural – usually adopt a paternalistic view and evaluate a policy intervention in terms of its *uncompensated* effects on individual outcomes. But they ignore

*Address for correspondence: debobhatta@gmail.com. I am grateful to Pascaline Dupas, Phil Haile, Jerry Hausman, Whitney Newey and seminar participants at the University of Wisconsin for feedback and suggestions. All errors are mine. JEL codes: C14, C25, D12, D61, H22

the heterogeneity in welfare effects which require computation of compensated effects. For example, in an educational context, researchers typically evaluate the impact of a tuition subsidy via its net effect on college-enrolment (c.f. Ichimura and Taber, 2002, Kane, 2003). But they stop short of evaluating how much the subsidy is valued by the potential beneficiaries themselves, viz., the lump-sum income transfer that would result in the same individual utilities as the subsidy, and any resulting deadweight loss thereof.[1] The present paper develops methods for calculating such welfare effects as part of program evaluation studies, without requiring the researcher to make *any* restrictive assumptions on the nature of preference-heterogeneity or income effects. The setting is where we observe individual level data on a cross-section of consumers making choices among discrete prospects, and the goal is to estimate the effects of hypothetical economic changes on consumer welfare.

In practical settings involving cross-sectional data, unobservable heterogeneity in consumer preferences makes empirical welfare analysis a challenging problem. Nonetheless, some advances based on Roy's identity have recently been made for the case of a continuous good, such as gasoline consumption, under general preference heterogeneity (c.f., Hausman and Newey, 2016, Lewbel and Pendakur, 2016). However, many important real-life decisions involve discrete choice, such as college-attendance, choice of commuting method, school-choice, retirement decisions, and so forth. The Roy's identity based methods used for continuous choice cannot be applied in these settings owing to the non-smoothness of individual demand in price and income.[2] Until very recently, available methods for welfare analysis in discrete choice settings were based on restrictive and arbitrary assumptions on preference heterogeneity, e.g., quasi-linear preferences implying absence of income effects (c.f., Domencich and McFadden, 1975, Small and Rosen 1981), or parametrically specified utility functions and heterogeneity distributions (Herriges and Kling, 1999, Dagsvik and Karlstrom, 2005, Goolsbee and Klenow, 2006). See McFadden, 1981, for an early discussion of para-

[1] Stiglitz (2000, page 276) notes that empirical researchers typically ignore income effects owing to the perceived difficulty of calculating them, and Goolsbee (1999, page 10) points out that whereas the economic theory of welfare largely relates to compensated elasticities, common program evaluation studies in public finance typically report uncompensated effects. Hendren (2013) discusses this point and some related issues in more detail.

[2] The closest parallel to Roy's Identity for discrete choice is the so-called Daly-Williams-Zachary theorem (McFadden, 1981, 5.8), which shows that in an additive random utility model with scalar heterogeneity, the choice probabilities equal certain derivatives of the *average* indirect utility. This result is not useful for analysis of *individual* welfare distributions, since the income compensation that maintains average utility is not the same as the average of the income compensations that maintain individual utilities, unless preferences are quasilinear. We are interested in the latter distribution, and hence the DWZ theorem is not relevant to our problem.

metric approaches to modelling discrete choice and welfare analysis. Indeed, two key concerns with parametric analyses are (a) model mis-specification leading to erroneous substantive conclusions, and (b) identification of welfare distributions solely from functional form assumptions.

Recently, Bhattacharya, 2015 (DB15, henceforth) has shown that for heterogeneous consumers facing the choice between mutually exclusive discrete alternatives, the marginal distributions of equivalent/compensating variation (EV/CV) resulting from a ceteris paribus price-change of an alternative can be expressed as closed-form transformations of conditional choice probabilities. These results hold under fully unrestricted heterogeneity and income-effects across consumers. Taking DB15 as the point of departure, the present paper makes three new contributions. First, we show that in a discrete choice setting, money-metric welfare effects of simultaneous price changes of several alternatives continue to remain well-defined whereas the analog of the Marshallian Consumer Surplus becomes path-dependent (Proposition 1). Next, EV/CV distributions in this case are shown to be expressible as closed-form transformations of estimable choice-probabilities (Theorem 1). These results cover situations where some price changes are negative, some zero and some positive. The key issue here, elaborated in Section 2.1 below, is that although welfare effects of multiple price-changes are well-defined, their distributions *cannot* be obtained by iterating the single price-change result of DB15. This is because the income at which welfare distributions are to be evaluated varies in an unobservable way across individuals from the second iteration onwards, and thus cannot be conditioned on. Consequently, new results are required for welfare analysis in these situations.

Multiple price-changes are the likely consequences of a single initial price change of a product with substitutes. For example, a school-tuition subsidy in an area with child labor is likely to raise children's wages in response. The impact of such simultaneous price-changes on consumer welfare is usually *the* key consideration for policymakers in regard to their decision on whether to implement a proposed policy-change (c.f., Willig et al, 1991). It has been common practice in applications to use the so-called log-sum formula (c.f., Small and Rosen, 1981, Train, 2009) for welfare comparisons in these settings. This formula, though convenient, is based on strong, unsubstantiated assumptions like absence of income effects and extreme valued heterogeneity, and potentially leads to erroneous substantive conclusions (see our empirical example below for a concrete illustration of such errors).[3] Here we show that under completely general preference heterogeneity and income

[3] Some of these restrictive assumptions have subsequently been replaced with less stringent parametric assumptions, c.f. Dagsvik and Karlstrom, 2005, McFadden and Train, 2000, Herriges and Kling, 1999, Goolsbe and Klenow, 2006.

effects, one can express welfare-distributions resulting from multiple simultaneous price-changes in terms of choice probability functions, thereby reducing welfare-analysis to the problem of estimating (structural) choice probabilities. Crucially, our welfare-expressions (a) hold when income effects are non-negligible, as is likely for bigger purchases like children's education and consumer durables, and (b) apply to arbitrary patterns of price changes across alternatives, thereby making the results useful across a large range of empirical situations.

Section 3 of this paper concerns welfare loss resulting from elimination of an existing alternative (or, equivalently, welfare gain from having added a new alternative to the choice set), such as banning teenage wage-labour in a poor country. Previous econometric studies of such problems either ignored consumer heterogeneity, and implicitly assumed a "representative consumer" model, and/or worked under restrictive parametric assumptions (Willig et al, 1991, Hausman, 1996, Hausman and Leonard, 2002, Hausman, 2003, Trajtenberg, 1989).[4] In contrast, our set-up allows for (a) unrestricted consumer heterogeneity, and (b) arbitrary effects of eliminating the alternative on the price of *existing* alternatives, thereby making our results applicable to a wide variety of practical situations.[5] Importantly, these results are robust to failure of parametric assumptions, and clarify which features of welfare distributions can and which ones cannot be learnt from demand data alone without functional form assumptions. Indeed, our key result, stated below in Theorem 2, shows that Hicksian welfare distributions in this case can again be expressed as closed-form transformations of choice-probabilities without any assumption on heterogeneity. Nonetheless, calculating the tails of these distributions require observation of demand at prices where aggregate demand at an adjusted income level reaches zero. At high income levels, such prices may not be observed since producers have no incentive to set a price where aggregate demand for high income people is zero. In that case, a lower bound on the welfare distributions can be obtained using the value of aggregate demand at the highest observed price.[6] A corollary of our main result is that the heuristic empiri-

All of these still require specification of the dimension and distribution of unobserved heterogeneity and functional form of utility functions about which no a priori information is available.

[4] See Lewbel (2001) for an illuminating discussion of demand and welfare analysis with a representative consumer vis-a-vis allowing for preference heterogeneity.

[5] Welfare calculations in the two scenarios described above correspond to situations where the price vectors before and after are given. The process through which the final price vector following the relevant change is calculated requires modelling the supply side. See, Hausman and Leonard, 2002, page 256-8, for an illustration of the methodology.

[6] For common *parametric* models, like mixed logit, these issues are assumed away, and welfare distributions are point-identified simply via functional form assumptions.

cal practice of calculating welfare effects of new goods by integrating choice probabilities from the current price to infinity yields the average EV, if prices of substitutes remain unchanged; however, calculation of the average CV and/or allowing for prices of substitutes to change entail different expressions. These results can be used for "retrospective" calculation of welfare gains from having introduced a new alternative, simply by reversing the labels of CV/EV.

Section 4 of this paper considers multinomial choice with non-exclusive alternatives. For example, if a cable-TV company offering a sports package and a movie package raises product-prices, then the resulting welfare calculations requires new results because the packages are not exclusive alternatives for a potential consumer. We show that for multiple *non-exclusive* discrete goods, the welfare distribution resulting from a single price-change can be directly expressed as a closed-form functional of choice-probabilities but that for multiple price-changes cannot, *unlike* the case of multinomial choice among *exclusive* options. We then show how to construct nonparametric bounds for these distributions.

Taken together, these results provide new insights into welfare analysis in discrete choice situations, as well as providing practitioners with useful empirical tools for evaluating policy changes in real-life settings. In section 5, we discuss practical implementation of our results, and state a new and useful finding, viz. that under income endogeneity and corresponding to a price increase, the EV but not the CV can be used for legitimate welfare analysis even in the absence of instruments or control functions.

Proofs of all theoretical results and details of numerical calculations for the empirical illustration are provided in the appendix.

We end this section by emphasizing that our results establish the closed-form mapping between welfare distributions and structural choice probabilities, while imposing no restriction on preference heterogeneity or functional form of utilities. In other words, our results show that knowledge of welfare distributions is *exactly equivalent* to the knowledge of structural choice probabilities. Once the structural choice probabilities (or nonparametric bounds on them) are identified, our results deliver identification of the welfare distributions (or bounds on them) *regardless of whether price and/or income are endogenous.*[7]

[7]Indeed, if one allows for endogenous regressors, then the nonparametric point-identification of structural choice probabilities themselves is nontrivial, and would typically require control function type approaches (c.f., Blundell and Powell, 2003, Matzkin, 2008 and Berry and Haile, 2015). Alternatively, the results of Chesher (2005) can be used to derive nonparametric bounds on structural choice probabilities in presence of endogeneity. But in either case, one

2 Multiple Price-Changes

Set-up and notation: Consider a multinomial choice situation where alternatives are indexed by $j = 1, ..., J$; individual income is denoted by Y, and price of alternative j by P_j. Individual utility from choosing alternative j is $U_j(Y - P_j, \eta)$, $j = 1, ..., J$, where η denotes individual heterogeneity of unknown dimension; η is distributed in the population with unknown marginal CDF $F_\eta(\cdot)$. We have a cross-sectional random sample of consumers, and observe their characteristics including income, the prices they face, and the choice they make. The following analysis implicitly conditions on observable non-income characteristics.

Define the *structural* choice probability for alternative j evaluated at price vector $\mathbf{p}$ and income y, denoted $\{q_j(\mathbf{p}, y)\}$, $j = 1, ..., J$, as

$$q_j(\mathbf{p}, y) = \int 1\left\{U_j(y - p_j, r) > \max_{k \neq j}\{U_k(y - p_k, r)\}\right\} dF_\eta(r). \tag{1}$$

In words, if we randomly sample individuals from the population, and offer the price vector $\mathbf{p}$ and income y to each sampled individual, then a fraction $q_j(\mathbf{p}, y)$ will choose alternative j, in expectation.

Assumption 1 *Assume that for each η and for each $j = 1, ..., J$, the utility function $U_j(\cdot, \eta)$ is strictly increasing.*

Assumption 1 simply says that corresponding to making any choice, every consumer is strictly better off if they have more numeraire left in their pocket. Note that this assumption leaves the dimension of heterogeneity completely unspecified, and says nothing about how utility changes with unobserved heterogeneity.

Now consider a hypothetical change in the price vector from $\mathbf{p}_0 \equiv (p_{10}, p_{20}, ..., p_{J0})$ to $\mathbf{p}_1 \equiv (p_{11}, p_{21}, ..., p_{J1})$. Then the EV at income y for an η type consumer is the income reduction S in the initial situation that would lead to attainment of the eventual indirect utility. Formally, the EV is the solution S to the equation:

$$\begin{aligned} & \max\{U_1(y - p_{11}, \eta), U_2(y - p_{21}, \eta), ..., U_J(y - p_{J1}, \eta)\} \\ = \; & \max\{U_1(y - S - p_{10}, \eta), U_2(y - p_{20} - S, \eta), ..., U_J(y - p_{J0} - S, \eta)\}. \end{aligned} \tag{2}$$

can directly apply our results to the resulting choice probability estimates, and obtain corrsponding welfare estimates without requiring any additional modification due to endogeneity.

Similarly, the CV is the income compensation in the eventual situation necessary to restore the initial indirect utility; formally, the CV is the solution S to the equation:

$$\begin{aligned} & \max\left\{U_1\left(y+S-p_{11},\eta\right),U_2\left(y+S-p_{21},\eta\right),...,U_J\left(y+S-p_{J1},\eta\right)\right\} \\ = & \max\left\{U_1\left(y-p_{10},\eta\right),U_2\left(y-p_{20},\eta\right),...,U_J\left(y-p_{J0},\eta\right)\right\}. \end{aligned} \tag{3}$$

As η varies in the population, the CV and EV will have a distribution across consumers. Our goal is to estimate these distribution functions using the cross-sectional dataset.

Note that by analogy with the single price change case, one can attempt to define the change in average Marshallian consumer surplus in the multiple price change case via the line-integral

$$CS\left(L\right)=-\int_L\sum_{j=1}^J q_j\left(\mathbf{p},y\right)dp_j, \tag{4}$$

where L denotes a path from $\mathbf{p}_0$ to $\mathbf{p}_1$ (c.f. Auerbach, 1985, equation 2.2). The negative sign stems from the fact that rise in price leads to a loss in consumer surplus.

In the set-up described above, we first show that for arbitrary price changes, the EV and CV are well-defined under assumption 1, but the Marshallian Consumer Surplus is not. We then show why the results of DB15 cannot be iterated to get welfare-distributions for multiple price changes, and then establish the first key result of this paper, viz., that the marginal distributions of individual-level EV and CV for arbitrary changes in the price vector can be obtained as closed-form functionals of the structural choice probabilities.

We will assume without loss of generality that

$$p_{J1}-p_{J0}\geq p_{J-1,1}-p_{J-1,0}\geq...\geq p_{11}-p_{10}. \tag{5}$$

That is, label the alternative with the smallest price change (the smallest could be a negative number, representing a fall in price) alternative 1, the next smallest as alternative 2 and so on.

Now, it is well known that for continuous choice with multiple prices changing simultaneously, the Marshallian Consumer Surplus is generically undefined in the sense that the corresponding line-integral is path-dependent, but the Hicksian CV and EV continue to remain well-defined (c.f., Tirole, 1988, page 11). Our first result establishes that the same conclusion holds for discrete choice.

Proposition 1 *Consider the multinomial choice set-up with J alternatives. Consider a price change from $\mathbf{p}_0 \equiv (p_{10}, p_{20}, ..., p_{J0})$ to $\mathbf{p}_1 \equiv (p_{11}, p_{21}, ..., p_{J1})$. Under assumption 1, the individual compensating and equivalent variations are uniquely defined.*

(Proof in Appendix)

Marshallian Consumer Surplus: It can be shown that for discrete choice with *multiple* price changes, the integral (4) is path-dependent. In the appendix, we demonstrate this for the case with 3 alternatives ($J = 3$). Thus it is no longer true that the average Marshallian consumer surplus is identical to the average EV, as found in DB15 for the case of a ceteris paribus price increase of a *single* alternative.

2.1 Welfare Distributions for Multiple Price-Changes

DB15 showed that in the above setting, when the price of a single alternative changes ceteris paribus, the resulting CV and EV distributions can be expressed as closed-form functionals of choice-probabilities. It is important to note that one cannot iterate this single price-change result to obtain the welfare distribution for simultaneous changes in the prices of *multiple* alternatives. To see this, consider a choice among three alternatives ($J = 3$), and suppose that price of alternative 1 changes from p_{10} to p_{11}, and that of 2 from p_{20} to p_{21}, and price of 3 is unchanged at p_3. Suppose we try to calculate the overall CV, starting with, say, price change of alternative 1, followed by 2 (the order in which we do this does not matter, by path independence) and applying theorem 2 of DB15 at each stage. Suppose the CVs corresponding to the two price changes are denoted by S_1 and S_2 respectively. Then, by definition,

$$\begin{aligned} & \max \left\{ U_1 \left(y - p_{10}, \eta \right), U_2 \left(y - p_{20}, \eta \right), U_3 \left(y - p_3, \eta \right) \right\} \\ = \; & \max \left\{ U_1 \left(y + S_1 - p_{11}, \eta \right), U_2 \left(y + S_1 - p_{20}, \eta \right), U_3 \left(y + S_1 - p_3, \eta \right) \right\} \\ = \; & \max \left\{ \begin{array}{l} U_1 \left(y + \underbrace{S_1 + S_2}_{=S, \text{ overall CV}} - p_{11}, \eta \right), U_2 \left(y + \underbrace{S_1 + S_2}_{S} - p_{21}, \eta \right), \\ U_3 \left(y + \underbrace{S_1 + S_2}_{S} - p_3, \eta \right). \end{array} \right\} \end{aligned}$$

Then using theorem 2 of DB15, we can get the marginal distribution of S_1 but we cannot get the marginal distribution of $S_1 + S_2$ because the price of both alternative 1 and 2 have changed between lines 1 and 3 of the previous display and so theorem 2 of DB15 does not apply. Secondly, because we cannot calculate the *value* of the CV S_1 for an individual (we can only calculate its *distribution* across all individuals), we cannot apply theorem 2 of DB15 to calculate the marginal distribution of S_2, since the income at which we could potentially apply the theorem depends on S_1 which is unknown, unlike y which is a fixed known constant. Thus, a new result is required for welfare analysis corresponding to *simultaneous* changes in multiple prices, and it is given by the following theorem.

Theorem 1 *Consider the multinomial choice set-up with J exclusive alternatives. Consider a price change from $\mathbf{p}_0 \equiv (p_{10}, p_{20}, ..., p_{J0})$ to $\mathbf{p}_1 \equiv (p_{11}, p_{21}, ..., p_{J1})$ satisfying (5). Denote $p_{j1} - p_{j0}$ by Δp_j for $j = 1, ..., J$. Under assumption 1, the marginal distribution of the individual EV evaluated at income y is given by*

$$\Pr(EV \leq a) = \begin{cases} 0 \text{ if } a < \Delta p_1, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{11}, ..., p_{j1}, \\ p_{j+1,0} + a, ..., p_{J0} + a, y \end{pmatrix} \text{ if } \Delta p_j \leq a < \Delta p_{j+1},\ a \geq 0,\ 1 \leq j \leq J-1, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{11} - a, ..., p_{j1} - a, \\ p_{j+1,0}, ..., p_{J0}, y - a \end{pmatrix} \text{ if } \Delta p_j \leq a < \Delta p_{j+1},\ a < 0,\ 1 \leq j \leq J-1, \\ 1 \text{ if } a \geq \Delta p_J, \end{cases} \tag{6}$$

while that of the individual CV evaluated at income y is given by

$$\Pr(CV \leq a) = \begin{cases} 0 \text{ if } a < \Delta p_1, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{11}, ..., p_{j1}, \\ p_{j+1,0} + a, ..., p_{J0} + a, y + a \end{pmatrix} \text{ if } \Delta p_j \leq a < \Delta p_{j+1},\ a \geq 0,\ 1 \leq j \leq J-1, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{11} - a, ..., p_{j1} - a, \\ p_{j+1,0}, ..., p_{J0}, y \end{pmatrix} \text{ if } \Delta p_j \leq a < \Delta p_{j+1},\ a < 0,\ 1 \leq j \leq J-1, \\ 1 \text{ if } a \geq \Delta p_J, \end{cases} \tag{7}$$

where q_ks are defined above in equation (1). (The separate entries for $a \geq 0$ and $a < 0$ in each line of (6) and (7) arise from accommodating rise and fall of prices, respectively).

These distributional results cover positive, zero and negative price-changes. For example, in the 3-alternative case, suppose alternative 2 is the outside option with price $p_{21} = p_{20} = 0$, and

$\Delta p_1 < 0 < \Delta p_3$, then (7) becomes

$$\Pr(CV \leq a) = \begin{cases} 0 \text{ if } a < \Delta p_1, \\ q_1(p_{11} - a, 0, p_{30}, y) \text{ if } \Delta p_1 \leq a < 0, \\ q_1(p_{11}, 0, p_{30} + a, y + a) + q_2(p_{11}, 0, p_{30} + a, y + a) \text{ if } 0 \leq a < \Delta p_3, \\ 1 \text{ if } a \geq \Delta p_3. \end{cases} \tag{8}$$

Finally, note that in the above results, the CDFs are non-decreasing because of assumption 1. For instance, for the CV, we require

$$q_1(p_{11}, p_{20} + a, p_{30} + a, y + a) \leq q_1(p_{11}, p_{20} + a', p_{30} + a', y + a') \tag{9}$$

whenever $p_{11} - p_{10} \leq a < a'$. But this is true because the LHS is the probability of the event

$$\begin{aligned} & U_1(y + a - p_{11}, \eta) \geq \max\{U_2(y - p_{20}, \eta), U_3(y - p_{30}, \eta)\} \\ \Rightarrow \quad & U_1(y + a' - p_{11}, \eta) \geq \max\{U_2(y - p_{20}, \eta), U_3(y - p_{30}, \eta)\}, \text{ since } a' > a, \\ \Longleftrightarrow \quad & U_1(y + a' - p_{11}, \eta) \geq \max\left\{ \begin{array}{c} U_2(y + a' - (p_{20} + a'), \eta), \\ U_3(y + a' - (p_{30} + a'), \eta) \end{array} \right\}, \end{aligned}$$

whose probability is the RHS of the previous display. Inequality (9) can be interpreted as a Slutsky/Revealed Preference inequality for discrete choice.

Remark 1 *Note that the above results hold no matter whether price and income are endogenous or not. Endogeneity affects how the (structural) choice probabilities are to be consistently estimated, not the relationship between welfare distributions and the structural choice probabilities which is what the above results establish. See Section 5 below for further discussion of welfare estimation under endogeneity.*

Remark 2 *It is also implicit throughout that the price changes do not alter the population of interest, e.g. a large tuition subsidy in a district might attract outsiders with a strong preference for education to migrate in, altering the distribution of preferences relative to the status-quo. In other words, the price changes considered here are assumed to be modest enough to have no impact on the distribution of* η.

3 Two Extensions

In this section we show how to calculate welfare effects in two practically relevant scenarios – viz., (i) elimination/introduction of an alternative from/to consumers' choice sets, and (ii) price changes when alternatives are not exclusive.

3.1 Elimination of an Alternative

Consider a setting of multinomial choice among exclusive alternatives $\{1,...,J+1\}$. Suppose the alternative $J+1$ is eliminated subsequently, which can potentially affect consumer welfare by both restricting the choice set and also by affecting the prices of other alternatives. Assume that we have data on a cross-section of individual choices in the pre-elimination situation. We wish to calculate the distribution of Hicksian welfare effects that would result from eliminating the $J+1$th alternative. Previous researchers (c.f. Hausman, 2003) had provided convenient, off-the-shelf formulae for estimation of welfare effects for a "representative" consumer, ignoring the impact of unobserved heterogeneity in preferences. Incorporation of preference heterogeneity and developing the associated welfare analysis reveals interesting differences between CV and EV based formulae and identifiability of their distribution, as will be shown below.

Toward that end, consider an individual at income y and unobserved heterogeneity η whose utility from consuming alternative j at price p_j is given by $U_j(y-p_j,\eta)$. The problem is to find the distribution of welfare effects across such individuals resulting from potentially eliminating alternative $J+1$. Suppose from an initial price vector $(p_{11},...,p_{J1},p_{J+1})$, following elimination of the $J+1$th alternative, the eventual price vector becomes $(p_{10},...,p_{J0})$.

Then the CV is the income compensation in the post-elimination situation necessary to attain the initial indirect utility. Formally, CV is the solution S to the equation:

$$\begin{aligned}&\max\left\{U_1(y-p_{11},\eta),...,U_J(y-p_{J1},\eta),U_{J+1}(y-p_{J+1},\eta)\right\}\\=&\max\left\{U_1(y+S-p_{10},\eta),...,U_J(y+S-p_{J0},\eta)\right\}.\end{aligned}$$

Analogously, the EV is defined as the income reduction S in the initial situation that would lead to attainment of the eventual, i.e. post-elimination indirect utility:

$$\begin{aligned}&\max\left\{U_1(y-S-p_{11},\eta),...,U_J(y-S-p_{J1},\eta),U_{J+1}(y-S-p_{J+1},\eta)\right\}\\=&\max\left\{U_1(y-p_{10},\eta),...,U_J(y-p_{J0},\eta)\right\}.\end{aligned}$$

Define the pre-elimination choice probabilities for alternatives $k=1,...,J+1$, as

$$q_k(p_1,...,p_{J+1},y)=\Pr\left[U_k(y-p_k,\eta)>\max_{j\in\{1,...,J+1\}\setminus\{k\}}U_j(y-p_j,\eta)\right]. \quad (10)$$

Assume WLOG that $p_{J,0}-p_{J,1}\geq...\geq p_{10}-p_{11}$, and denote $p_{j0}-p_{j1}$ by Δp_j for $j=1,...,J$.

We now state the main result describing the marginal distribution of EV and CV. The proof appears in the appendix.

Theorem 2 *Assume that for each* $j = 1, ..., J+1$, $U_j(\cdot, \eta)$ *is strictly increasing. Let* $q_k(\cdot, ..., \cdot, y)$ *be as defined in (10). Then*

$$\Pr(CV \leq a) = \begin{cases} 0 \text{ if } a < \Delta p_1, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{10}, ..., p_{j0}, \\ p_{j+1,1} + a, ..., p_{J1} + a, p_{J+1} + a, \\ y + a \end{pmatrix}, \text{ if } \Delta p_j \leq a < \Delta p_{j+1},\ j = 1, .., J-1,\ a \geq 0, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{10} - a, ..., p_{j0} - a, \\ p_{j+1,1}, ..., p_{J1}, p_{J+1}, y \end{pmatrix}, \text{ if } \Delta p_j \leq a < \Delta p_{j+1},\ 1 \leq j < J-1,\ a < 0, \\ 1 - q_{J+1}(p_{10}, ..., p_{J0}, p_{J+1} + a, y + a), \text{ if } \Delta p_J \leq a,\ a \geq 0, \\ 1 - q_{J+1}(p_{10} - a, ..., p_{J0} - a, p_{J+1}, y), \text{ if } \Delta p_J \leq a < 0. \end{cases} \tag{11}$$

On the other hand,

$$\Pr(EV \leq a) = \begin{cases} 0 \text{ if } a < p_{10} - p_{11}, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{10}, ..., p_{j0}, \\ p_{j+1,1} + a, ..., p_{J1} + a, p_{J+1} + a, y \end{pmatrix}, \text{ if } \Delta p_j \leq a < \Delta p_{j+1},\ j = 1, .., J-1,\ a \geq 0, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{10} - a, ..., p_{j0} - a, \\ p_{j+1,1}, ..., p_{J1}, p_{J+1}, y - a \end{pmatrix}, \text{ if } \Delta p_j \leq a < \Delta p_{j+1},\ 1 \leq j < J-1,\ a < 0, \\ 1 - q_{J+1}(p_{10}, ..., p_{J0}, p_{J+1} + a, y), \text{ if } \Delta p_J \leq a,\ a \geq 0, \\ 1 - q_{J+1}(p_{10} - a, ..., p_{J0} - a, p_{J+1}, y - a), \text{ if } \Delta p_J \leq a < 0. \end{cases} \tag{12}$$

As in the previous theorem, the pairs of results for $a < 0$ and $a \geq 0$ correspond to which existing alternatives have become more and less expensive, respectively, following the elimination of the $J+1$th alternative. Also, note that in order to calculate the probabilities appearing in theorem 2, we need to observe adequate cross-sectional variation in the price of all $J+1$ alternatives in the pre-elimination period.

Remark 3 *The above theorem also provides a formal justification of the heuristic empirical practice of equating elimination of an alternative with increasing its price to plus infinity for the purpose of welfare analysis.*

Corollary 1 *If elimination of the alternative has no effect on prices of the other alternatives, then* $p_{j1} = p_{j0}$ *for all* $j = 1, ..., J$*, and the above results simplify to*

$$\Pr(CV \leq a) = \begin{cases} 0 \text{ if } a < 0, \\ 1 - q_{J+1}(p_{10}, ..., p_{J0}, p_{J+1} + a, y + a) \text{ if } 0 \leq a. \end{cases} \tag{13}$$

$$\Pr(EV \leq a) = \begin{cases} 0 \text{ if } a < 0, \\ 1 - q_{J+1}(p_{10}, ..., p_{J0}, p_{J+1} + a, y) \text{ if } 0 \leq a. \end{cases} \tag{14}$$

These corollaries have clear intuitive interpretation. For example, consider the result (13). Recall that CV is defined as the solution S to

$$\begin{aligned} & \max\{U_1(y - p_{10}, \eta), ..., U_J(y - p_{J0}, \eta), U_{J+1}(y - p_{J+1}, \eta)\} \\ = & \max\{U_1(y + S - p_{10}, \eta), ..., U_J(y + S - p_{J0}, \eta)\}. \end{aligned}$$

Observe that any individual who does not consume the $J+1$th product in the pre-elimination situation suffers no welfare change relative to the final situation when the product becomes unavailable. Therefore, we only need to consider those individuals who consume the $J+1$th alternative in the pre-elimination situation. Among these individuals, the CV is positive and less than a for those who, with the compensated income $y + a$, would enjoy a higher utility than what they are getting from consuming the $J+1$th alternative in the pre-elimination situation. Since buying the $J+1$th alternative at price p_{J+1} with income y yields the same utility as buying it at price $p_{J+1} + a$ with income $y + a$, the probability of CV being less than a equals the probability of buying one of $1, ..., J$ when the price vector is $(p_{10}, ...p_{J0}, p_{J+1} + a)$ and income is $y + a$.

Corollary 2 *If eliminating the* $J+1$*th alternative has no effect on prices of the other alternatives, then the average values of individual welfare change are given by*

$$E(CV) = \int_{p_{J+1}}^{\infty} q_{J+1}(p_{10}, ..., p_{J0}, r, y + r - p_{J+1})\, dr, \tag{15}$$

$$E(EV) = \int_{p_{J+1}}^{\infty} q_{J+1}(p_{10}, ..., p_{J0}, r, y)\, dr, \tag{16}$$

using the change of variable $r = p_{J+1} + a$.

The expression (16) is commonly used as an ad-hoc measure of the welfare effect of introducing a new product. Thus it follows from the above discussion that if elimination of the alternative entails no price change for the other alternatives, then the commonly used expression happens to equal the mean EV, but ceases to be so if one is interested in the mean CV, or prices of substitutes also change.

In order to calculate the above expressions *nonparametrically*, a researcher needs to observe demand up to the price where the choice probability becomes zero. Typically, in a dataset, one is unlikely to observe such prices, since producers have no incentive to raise prices where revenue is zero. But then one can obtain a lower bound for $E(CV)$ by integrating up to the highest price observed in the dataset where demand from consumers is non-zero. This is in contrast to the case of price changes for multiple alternatives, reported in equation (6) above, where welfare distributions are nonparametrically identified as long as the hypothetical price-changes are within the range of the observed price data. Of course, for *parametric* choice probabilities, e.g. random coefficient logit, expressions like (16) are identified directly from functional form assumptions.

Finally, note that the above expressions can also be used for retrospective calculation of welfare distributions corresponding to *introduction* of a new alternative, simply by interchanging the labels of EV and CV. For example, Hausman, 2003, calculates the welfare effects of introducing a new brand of breakfast cereals when one has consumption data on all cereal brands including the recently introduced one. Hausman's calculations used a representative consumer idea, ignoring preference heterogeneity in the population. In contrast, allowing for unrestricted preference heterogeneity would lead to expressions for CV and EV distributions given by (12) and (11), respectively.

3.2 Non-exclusive Discrete Choice

We now change the set-up described above to allow for non-exclusive choice. Accordingly, assume that there are two binary choices which are non-exclusive among themselves. For example, suppose choice 1 for a household is whether to subscribe to a sports package offered by a cable TV network which costs P_1 and choice 2 is whether to subscribe to a movie package which costs P_2. A household then has four exclusive options – $\{1\}, \{2\}, \{1,2\}, \{0\}$ (where $\{0\}$ denotes choosing none of the two packages) with respective utilities $U_1(Y - P_1, \eta)$, $U_2(Y - P_2, \eta)$, $U_{12}(Y - P_1 - P_2, \eta)$ and $U_0(Y, \eta)$, respectively.

Single Price Change: Consider the CV corresponding to a rise in the price of the sports package from p_{10} to p_{11} with the price of the movie package fixed at p_2. The CV evaluated at income $Y = y$ is the solution to the equation

$$\begin{aligned} & \max\left\{ \begin{array}{l} U_0\left(y+CV,\eta\right), U_1\left(y+CV-p_{11},\eta\right), \\ U_2\left(y+CV-p_2,\eta\right), U_{12}\left(y+CV-p_{11}-p_2,\eta\right) \end{array} \right\} \\ = \; & \max\left\{ \begin{array}{l} U_0\left(y,\eta\right), U_1\left(y-p_{10},\eta\right), \\ U_2\left(y-p_2,\eta\right), U_{12}\left(y-p_{10}-p_2,\eta\right) \end{array} \right\}. \end{aligned} \tag{17}$$

The marginal distribution of CV corresponding to this single price change is point-identified in this case. The explanation of this result is as follows. Group option $\{1\}$ and $\{1,2\}$ together (call it group A) and options $\{0\}$ and $\{2\}$ together and call it group B. Define

$$\begin{aligned} & \varepsilon \overset{def}{=} (p_2, \eta) \\ & V_A\left(y-p_1,\varepsilon\right) \overset{def}{=} \max\left\{U_1\left(y-p_1,\eta\right), U_{12}\left(y-p_1-p_2,\eta\right)\right\}, \\ & V_B\left(y,\varepsilon\right) \overset{def}{=} \max\left\{U_0\left(y,\eta\right), U_2\left(y-p_2,\eta\right)\right\}. \end{aligned}$$

Correspondingly (17) becomes

$$\max\left\{V_A\left(y+CV-p_{11},\varepsilon\right), V_B\left(y+CV,\varepsilon\right)\right\} = \max\left\{V_A\left(y-p_{10},\varepsilon\right), V_B\left(y,\varepsilon\right)\right\}. \tag{18}$$

If the U functions are strictly increasing in the first argument for each η, then so are $V_A\left(\cdot,\varepsilon\right)$ and $V_B\left(\cdot,\varepsilon\right)$ for each ε. Now we can apply theorem 1 of DB15 for binary choice to this problem and get the marginal distribution of the compensating variation. For example, for $0 \leq a < p_{11} - p_{10}$, theorem 1 of DB15 gives

$$\begin{aligned} & \Pr\left(CV \leq a\right) \\ = \; & \Pr\left[V_B\left(y+a,\varepsilon\right) \geq V_A\left(y+a-\left(p_{10}+a\right),\varepsilon\right)\right] \\ = \; & \Pr\left[\begin{array}{l} \max\left\{U_0\left(y+a,\eta\right), U_2\left(y+a-p_2,\eta\right)\right\} \\ \geq \max\left\{ \begin{array}{c} U_1\left(y+a-\left(p_{10}+a\right),\eta\right), \\ U_{12}\left(y+a-\left(p_{10}+a\right)-p_2,\eta\right) \end{array} \right\} \end{array} \right] \\ = \; & q_0\left(p_{10}+a, p_2, y+a\right) + q_2\left(p_{10}+a, p_2, y+a\right). \end{aligned}$$

Thus

$$\Pr\left(CV \leq a\right) = \begin{cases} 0 \text{ if } a < 0, \\ q_0\left(p_{10}+a, p_2, y+a\right) + q_2\left(p_{10}+a, p_2, y+a\right), \text{ if } 0 \leq a < p_{11}-p_{10}, \\ 1 \text{ if } a \geq p_{11}-p_{10}. \end{cases}$$

Multiple Price Changes: The key fact enabling us to write (17) as the binary choice CV (18) is that P_2 is being held fixed at p_2; if P_2 also varied across individuals, then the distribution of ε would vary beyond the variation of η and the binary formulation would no longer be applicable. Indeed, for multiple price changes in the non-exclusive alternatives case, welfare distributions can no longer be written in terms of choice probabilities. To see this, consider a simultaneous rise in P_1 and P_2 from (p_{10}, p_{20}) to (p_{11}, p_{21}). Assume that $0 < p_{11} - p_{10} < p_{21} - p_{20}$.

Then the CV is defined via

$$\begin{aligned} & \max\left\{\begin{array}{c} U_0\left(y+CV,\eta\right), U_1\left(y+CV-p_{11},\eta\right), \\ U_2\left(y+CV-p_{21},\eta\right), U_{12}\left(y+CV-p_{11}-p_{21},\eta\right) \end{array}\right\} \\ = \;& \max\left\{U_0\left(y,\eta\right), U_1\left(y-p_{10},\eta\right), U_2\left(y-p_{20},\eta\right), U_{12}\left(y-p_{10}-p_{20},\eta\right)\right\}. \end{aligned}$$

Now,

$$\begin{aligned} & \Pr\left[CV = p_{11}-p_{10}\right] \\ = \;& \Pr\left[U_1\left(y-p_{10},\eta\right) \geq \max\left\{\begin{array}{l} U_0\left(y,\eta\right), U_2\left(y-p_{20},\eta\right), U_{12}\left(y-p_{10}-p_{20},\eta\right), \\ U_0\left(y+p_{11}-p_{10},\eta\right), U_2\left(y+p_{11}-p_{10}-p_{21},\eta\right), \\ U_{12}\left(y-p_{10}-p_{21},\eta\right) \end{array}\right\}\right] \\ = \;& \Pr\left[U_1\left(y-p_{10},\eta\right) \geq \max\left\{\begin{array}{l} U_2\left(y-p_{20},\eta\right), U_{12}\left(y-p_{10}-p_{20},\eta\right), \\ U_0\left(y+p_{11}-p_{10},\eta\right) \end{array}\right\}\right]. \end{aligned} \tag{19}$$

This probability is not generically point-identified from the choice-probabilities. To see this, consider the following counter-example.

Counter-example: Consider a classic McFadden type utility specification:$U_j\left(a,\eta\right) = \beta_j a + \eta_j$, for $j = 0, 1, 2, 12$, and the ηs distributed standardized extreme-valued. Then because the regressors $(y, y-p_1, y-p_2, y+p_1-p_2)$ constitute a 3 dimensional subspace – the sum of the 1st and the 4th regressors equals the sum of the 2nd and the 3rd implying exact multicollinearity – the 4 coefficients $(\beta_0, \beta_1, \beta_2, \beta_{12})$ are not separately identified. A direct way to verify this is to note that the Hessian of the likelihood function is of the form $W \otimes XX'$, where $X \equiv (y, y-p_{10}, y-p_{20}, y-p_{10}-p_{20})'$ (c.f. Bohning, 1992, page 198-9), and so its expectation is of the form $E\left(W \otimes XX'\right)$. Given that the 4th component of X is an exact linear combination of the first 3, in each block $E\left(W_{ij}XX'\right)$ of the Kronecker product the 4th row is exactly the same linear combination of the first 3 rows. So it follows that $E\left(W \otimes XX'\right)$ is globally singular. Therefore, by Rothenberg, 1971 theorem 1, the parameter vector β is not even locally identified.

Now the expression for $\Pr[CV = p_{11} - p_{10}]$ in this case equals

$$
\begin{aligned}
& q_1\left(y + p_{11} - p_{10}, y - p_{10}, y - p_{20}, y - p_{10} - p_{20}\right) \\
= \; & \frac{\exp\left(\beta_1\left(y - p_{10}\right)\right)}{\left(\begin{array}{c} \exp\left(\beta_0\left(y + p_{11} - p_{10}\right)\right) + \exp\left(\beta_1\left(y - p_{10}\right)\right) \\ + \exp\left(\beta_2\left(y - p_{20}\right)\right) + \exp\left(\beta_{12}\left(y - p_{10} - p_{20}\right)\right) \end{array}\right)}. \qquad (20)
\end{aligned}
$$

Since $x \equiv \left(y + p_{11} - p_{10}, y - p_{10}, y - p_{20}, y - p_{10} - p_{20}\right)$ spans a 4 dimensional space (unless $p_{11} \neq p_{10}$), we cannot have that for some $\theta \neq \beta$,

$$
\begin{aligned}
& q_1\left(y + p_{11} - p_{10}, y - p_{10}, y - p_{20}, y - p_{10} - p_{20}; \beta\right) \\
= \; & q_1\left(y + p_{11} - p_{10}, y - p_{10}, y - p_{20}, y - p_{10} - p_{20}; \theta\right)
\end{aligned}
$$

for all $p_{10}, p_{11}, p_{20}, y$. To see this explicitly, consider the following thought experiment. Suppose at a specific value x of the regressors, $q_1(x, \beta) = q_1(x, \theta)$. Now holding p_{10}, p_{20}, y fixed, if we increase p_{11}, then the denominator of $q_1(\cdot, \beta)$ will increase more (less) than that of $q_1(\cdot, \theta)$ if $\beta_0 > \theta_0$ ($\beta_0 < \theta_0$), so that $q_1(\cdot, \beta) = q_1(\cdot, \theta)$ cannot continue to hold. Therefore we must have $\beta_0 = \theta_0$. Next, if we increase y, p_{10} and p_{20} by the same amount such that $y - p_{10}$, $y - p_{20}$ remain fixed, only the 4th term in the denominator of (20) will change, and by the same logic as above, maintaining $q_1(\cdot, \beta) = q_1(\cdot, \theta)$ would require $\beta_{12} = \theta_{12}$. Changing y and p_{10} by the same amount holding p_{11} and p_{20} fixed would deliver $\beta_2 = \theta_2$, and changing y while changing p_{20} by the same amount and changing p_{11} the same amount in the opposite direction while holding p_{10} fixed would deliver $\beta_1 = \theta_1$.

Thus although $\theta \neq \beta$ will produce the same choice probabilities for all potentially observable values of prices and income, the expression (19) is different with positive probability (w.r.t. the joint distribution of the regressors) if $\theta \neq \beta$. This implies that the CV distribution cannot be identified from the choice probabilities unless $p_{11} = p_{10}$, which is the single price change scenario.

Bounds for the CV distribution: One can nonetheless bound the probability (19) using Assumption 1. In particular, let

$$
\begin{aligned}
A &= \left\{\begin{array}{c} p_1^*, p_2^*, y^* : y^* - p_1^* \geq y - p_{10},\ y^* - p_2^* \leq y - p_{20}, \\ y^* - p_1^* - p_2^* \leq y - p_{10} - p_{20},\ y^* \leq y + p_{11} - p_{10} \end{array}\right\}, \\
B &= \left\{\begin{array}{c} p_1^*, p_2^*, y^* : y^* - p_1^* \leq y - p_{10},\ y^* - p_2^* \geq y - p_{20}, \\ y^* - p_1^* - p_2^* \geq y - p_{10} - p_{20},\ y^* \geq y + p_{11} - p_{10} \end{array}\right\}.
\end{aligned}
$$

Then an upper bound consistent with Assumption 1 is given by:

$$\begin{aligned}
&\Pr\left[U_1\left(y-p_{10},\eta\right)\geq\max\left\{\begin{array}{l}U_2\left(y-p_{20},\eta\right),U_{12}\left(y-p_{10}-p_{20},\eta\right),\\ U_0\left(y+p_{11}-p_{10},\eta\right)\end{array}\right\}\right]\\
\leq\;&\inf_{\left(p_1^*,\,p_2^*,\,y^*\right)\in A}\Pr\left[U_1\left(y^*-p_1^*,\eta\right)\geq\max\left\{\begin{array}{l}U_2\left(y^*-p_2^*,\eta\right),U_{12}\left(y^*-p_1^*-p_2^*,\eta\right),\\ U_0\left(y^*,\eta\right)\end{array}\right\}\right]\\
=\;&\inf_{\left(p_1^*,\,p_2^*,\,y^*\right)\in A}q_1\left(y^*,y^*-p_1^*,y^*-p_2^*,y^*-p_1^*-p_2^*\right).
\end{aligned}\tag{21}$$

These bounds arise from the fact that if, e.g., $y^*-p_1^*\geq y-p_{10}$, then $U_1\left(y^*-p_1^*,\eta\right)\geq U_1\left(y-p_{10},\eta\right)$, by assumption 1, so that the probability of $U_1\left(y-p_{10},\eta\right)$ exceeding a specific number is smaller than that of $U_1\left(y^*-p_1^*,\eta\right)$ exceeding that same number. Similarly, $y^*-p_2^*\leq y-p_{20}$ implies that $U_2\left(y-p_{20},\eta\right)\geq U_2\left(y^*-p_2^*,\eta\right)$ so that the probability of $U_2\left(y-p_{20},\eta\right)$ being exceeded by a number is smaller than that of $U_2\left(y^*-p_2^*,\eta\right)$ being exceeded by it, etc.

By a similar logic, a lower bound on (19) is given by

$$\begin{aligned}
&\Pr\left[U_1\left(y-p_{10},\eta\right)\geq\max\left\{\begin{array}{l}U_2\left(y-p_{20},\eta\right),U_{12}\left(y-p_{10}-p_{20},\eta\right),\\ U_0\left(y+p_{11}-p_{10},\eta\right)\end{array}\right\}\right]\\
\geq\;&\sup_{\left(p_1^*,\,p_2^*,\,y^*\right)\in B}\Pr\left[U_1\left(y^*-p_1^*,\eta\right)\geq\max\left\{\begin{array}{l}U_2\left(y^*-p_2^*,\eta\right),U_{12}\left(y^*-p_1^*-p_2^*,\eta\right),\\ U_0\left(y^*,\eta\right)\end{array}\right\}\right]\\
=\;&\sup_{\left(p_1^*,\,p_2^*,\,y^*\right)\in B}q_1\left(y^*,y^*-p_1^*,y^*-p_2^*,y^*-p_1^*-p_2^*\right).
\end{aligned}\tag{22}$$ [8]

The conclusion from the above discussion is that in a discrete choice setting when alternatives are *non-exclusive*, the distributions of individual welfare change for a *single* ceteris paribus price change can be expressed directly as a choice probability, but for *simultaneous* change in several alternatives' prices, they can only be bounded by functionals of choice probabilities. This is in contrast to the multinomial case with *exclusive* alternatives where welfare distributions are identified for *both* single and multiple price changes. An intuition for the result is that in the non-exclusive case, there is a systematic relationship between the prices of the different options, so that in a dataset we can

[8] A topic of future research is to verify how tight these bounds would be in real applications, and how much they would shrink if we make additional behavioral assumptions, e.g. that one or both alternatives are normal goods, or that alternatives 1 and 2 are substitutes (or complements) for all consumers, etc. An obvious way to incorporate such assumptions would be to impose shape restrictions on the choice probabilities themselves, e.g. $\frac{\partial}{\partial y}\left(q_1\left(p_1,p_2,y\right)+q_{12}\left(p_1,p_2,y\right)\right)\geq 0$, or $\frac{\partial}{\partial p_2}\left(q_1\left(p_1,p_2,y\right)+q_{12}\left(p_1,p_2,y\right)\right)\geq 0$, respectively.

never observe independent price variations across the various (composite) options, no matter how many price combinations are observed across the individual alternatives.

Remark 4 *Note that this distinction between the exclusive and non-exclusive options would not be apparent if one started with a fully parametric model of choice; indeed, welfare distributions would appear to be expressible in terms of choice probabilities in both cases for arbitrary price-changes. This conclusion would, of course, arise solely from functional form assumptions.*

4 Discussions on Implementation

The results reported in theorems 1 and 2, and the associated corollaries are fully nonparametric in that no functional form assumptions are required to derive them. When implementing these results in practical applications, one can therefore estimate conditional choice probabilities nonparametrically, e.g., using kernel or series regressions and use these estimates to calculate welfare distributions. For instance, recall the 3-alternative case, leading to (8), where we have, say, $p_{11} - p_{10} < 0 = p_{21} - p_{20} < p_{31} - p_{30}$. For any random variable X with CDF $F(\cdot)$ and finite mean, the expectation satisfies

$$E(X) = -\int_{-\infty}^{0} F(x)\,dx + \int_{0}^{\infty} (1 - F(x))\,dx;$$

thus the mean CV from (8) is given by

$$-\int_{p_{11}-p_{10}}^{0} q_1(p_{11} - a, p_{20}, p_{30}, y)\,da + \int_{0}^{p_{31}-p_{30}} q_3(p_{11}, p_{20}, p_{30} + a, y + a)\,da, \tag{23}$$

where $q_j(p_1, p_2, p_3, y)$ denotes the choice probability of alternative j when the price of the three alternatives are (p_1, p_2, p_3) and income is y. If the dataset is of modest size, so that kernel regressions are imprecise, then one can alternatively use a parametric approximation to the choice probabilities. Numerical integration routines are now available in popular software packages like STATA and MATLAB, and can be used to calculate the integrals of choice probabilities in the same way that consumer surplus was traditionally calculated by earlier researchers.

4.1 Endogeneity

Income Endogeneity: In applications, observed income may be endogenous with respect to individual choice. A natural example is when omitted variables, such as unrecorded education level, can both determine individual choice and be correlated with income. Under such endogeneity, the *observed* choice probabilities would potentially differ from the *structural* choice probabilities, and one can define welfare distributions either unconditionally, or conditionally on income, analogous to the average treatment effect and the average effect of treatment on the treated, respectively, in the program evaluation literature. In this context, an important and useful insight, not previously noted, is that for a price-rise, the distribution of the income-conditioned EV is not affected by income endogeneity, whereas that of the CV is (for a fall in price, the conclusion is reversed for CV and EV respectively).

To see why that is the case, recall the three alternative case discussed above, and define the conditional-on-income structural choice probabilities as

$$q_j^c\left(p_1, p_2, p_3, y', y\right) = \int 1\left\{U_j\left(y' - p_j, \eta\right) \geq \max_{k \neq j} U_k\left(y' - p_k, \eta\right)\right\} dF\left(\eta | y\right),$$

where $F\left(\cdot | y\right)$ denotes the distribution of the unobserved heterogeneity η for individuals whose current income is y. Now, for a real number a, satisfying $p_{11} - p_{10} \leq a < p_{21} - p_{20}$, it is easy to see that similar to equation (8), the distributions of EV at a, evaluated at income y, conditional on current income being y, are given by

$$\Pr\left(EV \leq a | Inc = y\right) = q_1^c\left(p_{11}, p_{20} + a, p_{30} + a, y, y\right),$$

while for CV it is given by

$$\Pr\left(CV \leq a | Inc = y\right) = q_1^c\left(p_{11}, p_{20} + a, p_{30} + a, y + a, y\right).$$

Now, $q_1^c\left(p_{11}, p_{20} + a, p_{30} + a, y, y\right)$, by definition, is the fraction of individuals currently at income y who would choose alternative 1 at prices $\left(p_{11}, p_{20} + a, p_{30} + a\right)$, had their income been y. But this is directly observable in the data since the realized and hypothetical incomes are the same, and therefore, no corrections are required owing to endogeneity. However, $q_1^c\left(p_{11}, p_{20} + a, p_{30} + a, y + a, y\right)$ is the fraction of individuals currently at income y who would choose alternative 1 at prices $\left(p_{11}, p_{20} + a, p_{30} + a\right)$, *had their income been* $y + a$. This fraction is counterfactual and not directly estimable because the distribution of η is likely to be different across people with income $y + a$ relative to those with income y, due to endogeneity. To summarize, if the objective of welfare analysis is to calculate the EV distribution resulting from price rise for individuals whose realized income equals the hypothetical income, then endogeneity of income is irrelevant to the analysis. *This implies that if exogeneity of income is suspect and no obvious instrument or control func-*

tion is available, then a researcher can still perform meaningful welfare analysis based on the EV distribution at current income.

Price Endogeneity: When individual choice data are available, endogeneity of price is typically of lesser concern, because an individual's choice or her omitted characteristics are less likely to affect the market price she faces. However, if there are unobserved choice attributes, then price endogeneity may be an important empirical concern. In that case, the methods of Blundell and Powell using control functions can be used to identify structural choice probabilities. An alternative to the above is to obtain bounds for the average structural function using an instrument for price (c.f. Horowitz and Manski, Chesher), and then obtain bounds for the corresponding welfare distributions by integrating the choice probability bounds.

5 Summary and Conclusion

In this paper, we have shown how to conduct empirical welfare analysis in multinomial choice settings, allowing for completely general consumer heterogeneity and income-effects. The paper considers three scenarios – (a) simultaneous change in prices of multiple alternatives, (b) the introduction or elimination of an alternative, possibly accompanied by price-changes of other alternatives; and (c) situations where choice-alternatives are non-exclusive. The key results are that (i) Hicksian welfare changes are well-defined under a mild monotonicity assumption on utilities in all these cases, (ii) in cases (a) and (b) the marginal distributions of CV and EV can be expressed as simple closed form functions of choice probabilities without requiring any assumption on the functional forms of utilities, preference heterogeneity or income effects, and (iii) this last conclusion fails when alternatives are non-exclusive, but welfare distributions can still be bounded in that case. Our welfare measures are expressible as closed-form functionals of structural choice probabilities. As such, our approach strictly dominates state-of-the-art random-coefficient based demand and welfare analysis which require a researcher to either make arbitrary assumptions on the distribution of random coefficients (e.g. that they are normally distributed), or to require full support for regressors for identification and to solve difficult ill-posed inverse problems in estimation.

At a practical level, our methods can be used in program evaluation studies to calculate "compensated" program effects, i.e., the program's value to the subjects themselves, measured in terms of its cash equivalent, and the associated deadweight loss, without requiring restrictive assumptions on the nature of preference heterogeneity and income effects in the population. These money-metric welfare measures can also be compared across interventions with different outcomes. For example, a tuition subsidy for school attendance, and an adoption-subsidy for take-up of a health-product cannot be directly compared in terms of their average outcomes; but their cash equivalents (and the associated deadweight loss) can be directly compared, since both are expressed in monetary units.

References

1. Auerbach, Alan J. (1985): The theory of excess burden and optimal taxation. Handbook of public economics, vol 1, 61-127.

2. Berry, S. and P. Haile (2015). Identification in Differentiated Products Markets, NBER Working Paper No. 21500.

3. Bhattacharya, D. (2009). Inferring Optimal Peer Assignment from Experimental Data. Journal of the American Statistical Association 104, pp. 486-500.

4. Bhattacharya, D. (2015). Nonparametric Welfare Analysis for Discrete Choice, Econometrica, 83(2), 617–649.

5. Blundell, R. and J. Powell (2003). Endogeneity in Nonparametric and Semiparametric Regression Models, in Advances in Economics and Econometrics, Cambridge University Press, Cambridge, U.K.

6. Böhning, D. (1992). Multinomial logistic regression algorithm. Annals of the Institute of Statistical Mathematics, 44(1), pp.197-200.

7. 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.

8. Domencich, T. and D. McFadden (1975). Urban Travel Demand - A Behavioral Analysis, North-Holland, Oxford.

9. Goolsbee, A. 1999. Evidence on the High-Income Laffer Curve from Six Decades of Tax Reform. Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 30(2), pages 1-64.

10. Goolsbee, A. and P. Klenow. Valuing Consumer Products By The Time Spend Using Them: An Application To The Internet. American Economic Review, vol. 96, 108-113.

11. Hausman, J. (1981). Exact Consumer's Surplus and Deadweight Loss, The American Economic Review, Vol. 71, 4, 662-676.

12. Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition, The Economics of New Goods. University of Chicago Press, 207-248.

13. Hausman, J. and T. Leonard (2002). The Competitive Effects of a New Product Introduction: A Case Study, The Journal of Industrial Economics, Vol. 50 (3), pages 237–263.

14. Hausman, J. (2003). Sources of Bias and Solutions to Bias in the Consumer Price Index, The Journal of Economic Perspectives, Vol. 17, No. 1, 23-44.

15. Hausman, J. and Newey, W. (2015). Individual Heterogeneity and Average Welfare, CEMMAP working paper, 42/14, forthcoming, Econometrica.

16. Hendren, N. (2013). The Policy Elasticity. NBER Working Paper #19177

17. Herriges, J and C. Kling (1999). Nonlinear Income Effects in Random Utility Models, The Review of Economics and Statistics, Vol. 81, No. 1, 62-72.

18. Hicks, J. (1946). Value and Capital, Clarendon Press, Oxford.

19. Ichimura, H., and C. Taber (2002). Semiparametric reduced-form estimation of tuition subsidies. American Economic Review (2002): 286-292.

20. Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. The Economic Journal, pp. 549-552.

21. Kane, T. J. (2003). A quasi-experimental estimate of the impact of financial aid on college-going. No. w9703. National Bureau of Economic Research.

22. Kitamura, Y. and J. Stoye (2013). Nonparametric Analysis of Random Utility Models. Testing, mimeo. Cornell and Yale University.

23. Lewbel, A. (2001). Demand Systems with and without Errors, The American Economic Review, Vol. 91, No. 3.

24. Lewbel. A. and K. Pendakur (2015). Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients, mimeo. Boston College.

25. McFadden, D. (1981): Econometric Models of Probabilistic Choice, Manski, C. F., & McFadden, D. (Eds.) Structural analysis of discrete data with econometric applications, Chapter 5. Cambridge, MA: MIT Press.

26. Matzkin, R (2008): Identification in Nonparametric Simultaneous Equations, Econometrica, 76, pp 945–978.

27. D. McFadden, K. Train (2000). Mixed MNL models for discrete response. Journal of Applied Econometrics 15 (5), 447-470.

28. Newey, W. (1994). Kernel Estimation of Partial Means and a General Variance Estimator. Econometric Theory, 10, 233-253.

29. Rothenberg, T.J., 1971. Identification in parametric models. Econometrica: Journal of the Econometric Society, pp.577-591.

30. Small, K. and Rosen, H. (1981). Applied Welfare Economics with Discrete Choice Models. Econometrica. Vol. 49, No. 1, 105-130.

31. Stiglitz, J (2000). Economics of the public sector, W. W. Norton.

32. Train, K. (2009). Discrete Choice Methods with Simulation, Cambridge University Press, Cambridge, UK.

33. Trajtenberg, M. (1989): The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners. Journal of Political Economy, 97(2), 444-479.

34. Willig, R., S. Salop, and F. Scherer (1991). Merger analysis, industrial organization theory, and merger guidelines. Brookings Papers on Economic Activity. Microeconomics 281-332.

Appendix

Path-dependence of Line Integral Defining Marshallian Consumer Surplus for Discrete Choice

Consider a setting with three mutually exclusive alternatives with initial prices $\mathbf{p}_0 \equiv (p_{10}, p_{20}, p_{30})$ and final prices $\mathbf{p}_1 \equiv (p_{11}, p_{22}, p_{33})$. Let y denote income and $q_j(\mathbf{p}, y)$ denote the choice probability of alternative j when the price vector is $\mathbf{p}$ and income is y. Then the change in average consumer surplus arising from the price change from $\mathbf{p}_0$ to $\mathbf{p}_1$ can be defined via the line integral

$$CS(L) = -\int_L q_1(\mathbf{p}, y)\, dp_1 + q_2(\mathbf{p}, y)\, dp_2 + q_3(\mathbf{p}, y)\, dp_3,$$

where L denotes a path $L(t)$ from $t = 0$ to $t = 1$ such that $L(0) \equiv \mathbf{p}_0 \equiv (p_{10}, p_{20}, p_{30})$ and $L(1) \equiv \mathbf{p}_1 \equiv (p_{11}, p_{22}, p_{33})$. Consider two different such paths

$$\begin{aligned} L_1(t) &= (p_{10} + t(p_{11} - p_{10}), p_{20} + t(p_{21} - p_{20}), p_{30} + t(p_{31} - p_{30})) \\ L_2(t) &= \left(p_{10} + t^2(p_{11} - p_{10}), p_{20} + t(p_{21} - p_{20}), p_{30} + t(p_{31} - p_{30})\right). \end{aligned}$$

Then

$$CS(L_1) = -\int_0^1 \begin{bmatrix} (p_{11} - p_{10}) \times q_1(\mathbf{p}_0 + t(\mathbf{p}_1 - \mathbf{p}_0), y) \\ +(p_{21} - p_{20}) \times q_2(\mathbf{p}_0 + t(\mathbf{p}_1 - \mathbf{p}_0), y) \\ +(p_{31} - p_{30}) \times q_3(\mathbf{p}_0 + t(\mathbf{p}_1 - \mathbf{p}_0), y) \end{bmatrix} dt.$$

But

$$\begin{aligned} & CS(L_2) \\ = & -\int_0^1 \begin{bmatrix} 2t(p_{11} - p_{10}) \times q_1\left(\mathbf{p}_0 + t(\mathbf{p}_1 - \mathbf{p}_0) + (t^2 - t)(p_{11} - p_{10}), y\right) \\ +(p_{21} - p_{20}) \times q_2\left(\mathbf{p}_0 + t(\mathbf{p}_1 - \mathbf{p}_0) + (t^2 - t)(p_{11} - p_{10}), y\right) \\ +(p_{31} - p_{30}) \times q_3\left(\mathbf{p}_0 + t(\mathbf{p}_1 - \mathbf{p}_0) + (t^2 - t)(p_{11} - p_{10}), y\right) \end{bmatrix} dt, \end{aligned}$$

which would in general differ from $CS(L_1)$. Thus the CS is not well-defined for simultaneous change in multiple prices. Note that if only p_1 changes, then

$$\begin{aligned} CS(L_1) &= -(p_{11} - p_{10}) \times \int_0^1 [q_1(p_{10} + t(p_{11} - p_{10}), y)]\, dt, \\ CS(L_2) &= -2\int_0^1 \left[(p_{11} - p_{10}) \times q_1\left(p_{10} + t^2(p_{11} - p_{10}), y\right)\right] t dt \\ &= -(p_{11} - p_{10}) \times \int_0^1 q_1(p_{10} + r(p_{11} - p_{10}), y)\, dr, \text{ substituting } t^2 = r \\ &= CS(L_1), \end{aligned}$$

and we get back path independence. Thus the loss of path-independence arises only for multiple simultaneous price-changes.

Proof of Proposition 1

Proof. Let $\mathbf{p}_0 = (p_{10}, p_{20}, ..., p_{J0})$ denote the initial price vector and $\mathbf{p}_1 = (p_{11}, p_{21}, ..., p_{J1})$ denote the final price-vector. Suppose that two numbers S and T with $S \neq T$ solve the equation for the EV. Then, by definition,

$$\begin{aligned} & \max \{U_1 (y - p_{11}, \eta), ..., U_J (y - p_{J1}, \eta)\} \\ = & \max \{U_1 (y - p_{10} - S, \eta), ..., U_J (y - p_{J0} - S, \eta)\}, \end{aligned} \tag{24}$$

and

$$\begin{aligned} & \max \{U_1 (y - p_{11}, \eta), ..., U_J (y - p_{J1}, \eta)\} \\ = & \max \{U_1 (y - p_{10} - T, \eta), ..., U_J (y - p_{J0} - T, \eta)\}, \end{aligned} \tag{25}$$

so that

$$\begin{aligned} & \max \{U_1 (y - p_{10} - T, \eta), ..., U_J (y - p_{J0} - T, \eta)\} \\ = & \max \{U_1 (y - p_{10} - S, \eta), ..., U_J (y - p_{J0} - S, \eta)\}. \end{aligned} \tag{26}$$

Since each $U_j(\cdot)$, $j = 1, ...J$, is strictly increasing by assumption 2 if $S > T$, each term within $\{\}$ on the RHS of (26) will be strictly smaller than the corresponding term on the RHS. Therefore, each term within $\{\}$ on the RHS will be strictly smaller than the maximum value in the LHS. Since there are finitely many terms on the RHS, the maximum will also be strictly smaller than the maximum on the LHS, a contradiction. Similarly, if $S < T$, then the RHS of (26) will be strictly larger than the LHS. Therefore in order for (26) to hold, we must have $S = T$.

An exactly analogous argument works for the CV where the analogous equalities are

$$\begin{aligned} & \max \{U_1 (y - p_{11} + S, \eta), ..., U_J (y - p_{J1} + S, \eta)\} \\ = & \max \{U_1 (y - p_{10}, \eta), ..., U_J (y - p_{J0}, \eta)\} \\ = & \max \{U_1 (y - p_{11} + T, \eta), ..., U_J (y - p_{J1} + T, \eta)\}. \end{aligned}$$

Again, by assumption 1, this implies $S = T$. ■

Proof of Theorem 1. First, consider EV.

Note that the EV is defined by

$$\begin{aligned} & \max\left\{U_1\left(y-p_{11},\eta\right),U_2\left(y-p_{21},\eta\right),...,U_J\left(y-p_{J1},\eta\right)\right\} \\ = & \max\left\{U_1\left(y-S-p_{10},\eta\right),U_2\left(y-p_{20}-S,\eta\right),...,U_J\left(y-p_{J0}-S,\eta\right)\right\}. \end{aligned} \tag{27}$$

The first step is to establish that

$$EV \leq a \Leftrightarrow \left\{ \begin{array}{c} \max\left\{U_1\left(y-p_{11},\eta\right),U_2\left(y-p_{21},\eta\right),...,U_J\left(y-p_{J1},\eta\right)\right\} \\ \geq \max\left\{U_1\left(y-a-p_{10},\eta\right),U_2\left(y-p_{20}-a,\eta\right),...,U_J\left(y-p_{J0}-a,\eta\right)\right\} \end{array} \right\}. \tag{28}$$

Indeed, it is obvious from (27) that $EV \leq a$ will imply the RHS inequality inside the $\{\cdot\}$ in (28). To see the converse, assume that

$$\begin{aligned} & \max\left\{U_1\left(y-p_{11},\eta\right),U_2\left(y-p_{21},\eta\right),...,U_J\left(y-p_{J1},\eta\right)\right\} \\ \geq & \max\left\{U_1\left(y-a-p_{10},\eta\right),U_2\left(y-p_{20}-a,\eta\right),...,U_J\left(y-p_{J0}-a,\eta\right)\right\}. \end{aligned} \tag{29}$$

Now (29) and equation (27) imply that

$$\begin{aligned} & \max\left\{U_1\left(y-S-p_{10},\eta\right),U_2\left(y-p_{20}-S,\eta\right),...,U_J\left(y-p_{J0}-S,\eta\right)\right\} \\ \geq & \max\left\{U_1\left(y-a-p_{10},\eta\right),U_2\left(y-p_{20}-a,\eta\right),...,U_J\left(y-p_{J0}-a,\eta\right)\right\}, \end{aligned}$$

i.e., for all $j=1,...,J$, we have that

$$\max\left\{U_1\left(y-S-p_{10},\eta\right),...,U_J\left(y-p_{J0}-S,\eta\right)\right\} \geq U_j\left(y-p_{j0}-a,\eta\right).$$

If the maximum on the LHS of the previous display is the kth term, i.e., $U_k\left(y-S-p_{k0},\eta\right)$, then choosing $j=k$ on the RHS, we have that

$$U_k\left(y-S-p_{k0},\eta\right) \geq U_k\left(y-p_{k0}-a,\eta\right),$$

whence, applying assumption 1, it follows that $S \leq a$. This establishes (28).

Now we work from (28) to derive the CDF of the EV. To do this, first note that if $p_{l1}-p_{l0} \leq a < p_{l+1,1}-p_{l+1,0}$, $l=1,...,J-1$, then the inequality (29) can hold if and only if the LHS max is one of $U_1\left(y-p_{11},\eta\right),U_2\left(y-p_{21},\eta\right),...,U_l\left(y-p_{l1},\eta\right)$ but not $U_{l+1}\left(y-p_{l+1,1},\eta\right),...,U_J\left(y-p_{J1},\eta\right)$. To see this, suppose to the contrary that the max on the LHS of (29) is obtained for some k satisfying $J \geq k > l$. Then (29) implies

$$U_k\left(y-p_{k1},\eta\right) \geq \max_{j=1,...J}\left\{U_j\left(y-a-p_{j0},\eta\right)\right\} \geq U_k\left(y-a-p_{k0},\eta\right).$$

Given assumption 1, i.e., monotonicity of $U_k(\cdot, \eta)$, it follows that $p_{k1} \le a + p_{k0}$, i.e., $a \ge p_{k1} - p_{k0}$, a contradiction since $k > l$.

Therefore, if a satisfies $p_{l1} - p_{l0} \le a < p_{l+1,1} - p_{l+1,0}$, then

$$
\begin{aligned}
& \Pr(EV \le a) \\
= & \Pr\left[\begin{array}{c} U_1(y - p_{11}, \eta) \\ \ge \max \left\{ \begin{array}{c} U_2(y - p_{21}, \eta), ..., U_l(y - p_{l1}, \eta), ..., U_J(y - p_{J1}, \eta), \\ U_1(y - a - p_{10}, \eta), U_2(y - p_{20} - a, \eta), ..., \\ U_l(y - a - p_{l0}, \eta), ..., U_J(y - p_{J0} - a, \eta) \end{array} \right\} \end{array}\right] \\
& + \Pr\left[\begin{array}{c} U_2(y - p_{21}, \eta) \\ \ge \max \left\{ \begin{array}{c} U_1(y - p_{11}, \eta), ..., U_l(y - p_{l1}, \eta), ..., U_J(y - p_{J1}, \eta) \\ U_1(y - a - p_{10}, \eta), U_2(y - p_{20} - a, \eta), ..., \\ U_l(y - a - p_{l0}, \eta), ..., U_J(y - p_{J0} - a, \eta) \end{array} \right\} \end{array}\right] \\
& + ... \\
& + \Pr\left[\begin{array}{c} U_l(y - p_{l1}, \eta) \\ \ge \max \left\{ \begin{array}{c} U_1(y - p_{11}, \eta), U_2(y - p_{21}, \eta), ..., U_{l-1}(y - p_{l-1,1}, \eta), \\ U_{l+1}(y - p_{l+1,1}, \eta), ..., U_J(y - p_{J1}, \eta) \\ U_1(y - a - p_{10}, \eta), U_2(y - p_{20} - a, \eta), ..., \\ U_l(y - a - p_{l0}, \eta), ..., U_J(y - p_{J0} - a, \eta) \end{array} \right\} \end{array}\right] \\
\overset{(1)}{=} & \Pr\left[\begin{array}{c} U_1(y - p_{11}, \eta) \\ \ge \max \left\{ \begin{array}{c} U_2(y - p_{21}, \eta), ..., U_l(y - p_{l1}, \eta), \\ U_{l+1}(y - a - p_{l+1,0}), ..., U_J(y - p_{J0} - a, \eta) \end{array} \right\} \end{array}\right] \\
& + \Pr\left[\begin{array}{c} U_2(y - p_{21}, \eta) \\ \ge \max \left\{ \begin{array}{c} U_1(y - p_{11}, \eta), ..., U_l(y - p_{l1}, \eta), \\ U_{l+1}(y - a - p_{l+1,0}), ..., U_J(y - p_{J0} - a, \eta) \end{array} \right\} \end{array}\right] \\
& + ... \\
& + \Pr\left[\begin{array}{c} U_l(y - p_{l1}, \eta) \\ \ge \max \left\{ \begin{array}{c} U_1(y - p_{11}, \eta), ..., U_{l-1}(y - p_{l-1,1}, \eta), \\ U_{l+1}(y - a - p_{l+1,0}, \eta), ..., U_J(y - p_{J0} - a, \eta) \end{array} \right\} \end{array}\right]
\end{aligned}
$$

where the equality marked (1) uses the fact that $y - p_{j1} \ge y - a - p_{j,0}$, for all $j = 1, ..., l$, since

$p_{l1} - p_{l0} \leq a$. The above probability equals

$$\begin{aligned} & q_1 \left(p_{11}, p_{21}, ..., p_{l1}, p_{l+1,0} + a, ..., p_{J0} + a, y \right) \\ & + q_2 \left(p_{11}, p_{21}, ..., p_{l1}, p_{l+1,0} + a, ..., p_{J0} + a, y \right) \\ & ... + q_l \left(p_{11}, p_{21}, ..., p_{l1}, p_{l+1,0} + a, ..., p_{J0} + a, y \right), \end{aligned}$$

if $a \geq 0$, and equals

$$\begin{aligned} = \quad & q_1 \left(p_{11} - a, p_{21} - a, ..., p_{l1} - a, p_{l+1,0}, ..., p_{J0}, y - a \right) \\ & + q_2 \left(p_{11} - a, p_{21} - a, ..., p_{l1} - a, p_{l+1,0}, ..., p_{J0}, y - a \right) \\ & ... + q_l \left(p_{11} - a, p_{21} - a, ..., p_{l1} - a, p_{l+1,0}, ..., p_{J0}, y - a \right), \end{aligned}$$

if $a < 0$. This is precisely expression (6).

Next consider CV, defined as the solution S to the equation

$$\begin{aligned} & \max \left\{ U_1 \left(y + S - p_{11}, \eta \right), U_2 \left(y + S - p_{21}, \eta \right), ..., U_J \left(y + S - p_{J1}, \eta \right) \right\} \\ = \quad & \max \left\{ U_1 \left(y - p_{10}, \eta \right), U_2 \left(y - p_{20}, \eta \right), ..., U_J \left(y - p_{J0}, \eta \right) \right\}. \end{aligned} \tag{30}$$

The first step is to see that $CV \leq a$ is equivalent to

$$\begin{aligned} & \max \left\{ U_1 \left(y - p_{10}, \eta \right), U_2 \left(y - p_{20}, \eta \right), ..., U_J \left(y - p_{J0}, \eta \right) \right\} \\ \leq \quad & \max \left\{ U_1 \left(y + a - p_{11}, \eta \right), U_2 \left(y + a - p_{21}, \eta \right), ..., U_J \left(y + a - p_{J1}, \eta \right) \right\}. \end{aligned} \tag{31}$$

The necessity is obvious by assumption 1. Sufficiency follows because (30) and (31) imply that

$$\begin{aligned} & \max \left\{ U_1 \left(y + S - p_{11}, \eta \right), U_2 \left(y + S - p_{21}, \eta \right), ..., U_J \left(y + S - p_{J1}, \eta \right) \right\} \\ \leq \quad & \max \left\{ U_1 \left(y + a - p_{11}, \eta \right), U_2 \left(y + a - p_{21}, \eta \right), ..., U_J \left(y + a - p_{J1}, \eta \right) \right\}. \end{aligned} \tag{32}$$

Then there must be at least one j such that the RHS max is $U_j \left(y + a - p_{j1}, \eta \right)$, which, by (32) must be larger than $U_j \left(y + S - p_{j1}, \eta \right)$, implying by assumption 1 that $S \leq a$.

Next, suppose a satisfies $p_{j1} - p_{j0} \leq a < p_{j+1,1} - p_{j+1,0}$. Then $y + a - p_{j+1,1} < y - p_{j+1,0}$, and therefore, by assumption 1, $U_{j+1}(y + a - p_{j+1,1}, \eta) < U_{j+1} \left(y - p_{j+1,0}, \eta \right)$. Therefore, the RHS max in (31) must be one of $U_1(y+a-p_{11},\eta),...,U_j(y+a-p_{j1},\eta)$. Accordingly, we have that

$$
\begin{aligned}
&\Pr(CV\le a)\\
&=\sum_{k=1}^{j}\Pr\left[\begin{array}{c} U_k(y+a-p_{k1},\eta)\\ \ge\max\left\{\begin{array}{c} U_1(y-p_{10},\eta),...,U_{k-1}(y-p_{k-1,0},\eta),U_k(y-p_{k0},\eta),\\ U_{k+1}(y-p_{k+1,0},\eta)...,U_J(y-p_{J0},\eta)\\ U_1(y+a-p_{11},\eta),...,U_{j-1}(y+a-p_{j-1,1},\eta),\\ U_j(y+a-p_{j1},\eta),U_{j+1}(y+a-p_{j+1,1},\eta),\\ ...U_J(y+a-p_{J1},\eta)\end{array}\right\}\end{array}\right]\\
&=\sum_{k=1}^{j}\Pr\left[\begin{array}{c} U_k(y+a-p_{k1},\eta)\\ \ge\max\left\{\begin{array}{c} U_1(y-p_{10},\eta),...,U_{k-1}(y-p_{k-1,0},\eta),\\ U_{k+1}(y-p_{k+1,0},\eta)...,U_J(y-p_{J0},\eta)\\ U_1(y+a-p_{11},\eta),...,U_{j-1}(y+a-p_{j-1,1},\eta),\\ U_j(y+a-p_{j1},\eta),U_{j+1}(y+a-p_{j+1,1},\eta)\\ ...U_J(y+a-p_{J1},\eta)\end{array}\right\}\end{array}\right]\\
&=\sum_{k=1}^{j}\Pr\left[\begin{array}{c} U_k(y+a-p_{k1},\eta)\\ \ge\max\left\{\begin{array}{c} U_1(y+a-p_{11},\eta),...,U_j(y+a-p_{j1},\eta),\\ U_{j+1}(y-p_{j+1,0},\eta),...U_J(y-p_{J0},\eta)\end{array}\right\}\end{array}\right].
\end{aligned}
$$

The second equality follows from the fact that $k\le j$ and $p_{j1}-p_{j0}\le a$, and so we have by assumption 1 that $U_k(y+a-p_{k1},\eta)\ge U_k(y-p_{k0},\eta)$. The third equality follows from $p_{j1}-p_{j0}\le a<p_{j+1,1}-p_{j+1,0}$, so by assumption 1, $U_k(y-p_{k,0},\eta)\le U_k(y+a-p_{k1},\eta)$ for all $k\le j$ and $U_k(y-p_{k,0},\eta)>U_k(y+a-p_{k1},\eta)$ for all $k>j$. Finally,

$$
\begin{aligned}
&\sum_{k=1}^{j}\Pr\left[\begin{array}{c} U_k(y+a-p_{k1},\eta)\\ \ge\max\left\{\begin{array}{c} U_1(y+a-p_{11},\eta),...,U_j(y+a-p_{j1},\eta),\\ U_{j+1}(y-p_{j+1,0},\eta),...U_J(y-p_{J0},\eta)\end{array}\right\}\end{array}\right]\\
&\equiv\left\{\begin{array}{c} \sum_{k=1}^{j}q_k(p_{11},...,p_{j1},p_{j+1,0}+a,...,p_{J0}+a,y+a)\ \text{ if } a\ge 0,\\ \sum_{k=1}^{j}q_k(p_{11}-a,...,p_{j1}-a,p_{j+1,0},...,p_{J0},y)\ \text{ if } a<0.\end{array}\right.
\end{aligned}
$$

■

Proof of theorem 2:

Proof. The CV solves

$$\begin{aligned} & \max\left\{U_1\left(y-p_{11},\eta\right),...,U_J\left(y-p_{J1},\eta\right),U_{J+1}\left(y-p_{J+1},\eta\right)\right\} \\ = & \max\left\{U_1\left(y+S-p_{10},\eta\right),...,U_J\left(y+S-p_{J0},\eta\right)\right\}. \end{aligned}$$

Now, by the same logic as the one leading to equation (28) in the previous proof,

$$\begin{aligned} & \Pr\left[S\leq a\right] \\ = & \Pr\left[\begin{array}{c} \max\left\{U_1\left(y-p_{11},\eta\right),...,U_J\left(y-p_{J1},\eta\right),U_{J+1}\left(y-p_{J+1},\eta\right)\right\} \\ \leq \max\left\{U_1\left(y+a-p_{10},\eta\right),...,U_J\left(y+a-p_{J0},\eta\right)\right\} \end{array}\right]. \end{aligned} \tag{33}$$

Recall the assumption that (WLOG)

$$p_{J0}-p_{J1}\geq p_{J-1,0}-p_{J-1,1}\geq ...\geq p_{10}-p_{11}.$$

Therefore, if $a<p_{10}-p_{11}$, then for each $j=1,...,J$, the $U_j\left(y-p_{j1},\eta\right)$ on the LHS of (33) will be strictly larger than the corresponding $U_j\left(y+a-p_{j0},\eta\right)$, contradicting that the RHS max exceeds the LHS max. Thus there can be no probability mass below $p_{10}-p_{11}$. If for some $j\in\{1,...J-1\}$, $p_{j0}-p_{j1}\leq a<p_{j+1,0}-p_{j+1,1}$, then each of the first j terms on the RHS of (33) is at least as large as the corresponding term on the LHS, while the $(j+1)$th term onwards on the RHS are smaller than the corresponding terms on the RHS. This means that one of these first j terms on the RHS must be the maximum and it must also exceed the $j+1$th term onwards on the RHS. Thus, for $p_{j0}-p_{j1}\leq a<p_{j+1,0}-p_{j+1,1}$, the probability that $S\leq a$ equals

$$\begin{aligned} & \sum_{k=1}^{j}\Pr\left[U_k\left(y+a-p_{k0},\eta\right)\geq\max\left\{\begin{array}{c} \max_{k'\in\{1,...,j\}\backslash k}\left\{U_{k'}\left(y+a-p_{k'0},\eta\right)\right\}, \\ \max_{l\geq j+1}\left\{U_l\left(y-p_{l1},\eta\right)\right\}, \\ U_{J+1}\left(y-p_{J+1},\eta\right) \end{array}\right\}\right] \\ = & \left\{\begin{array}{c} \sum_{k=1}^{j}q_k\left(p_{10},...,p_{j0},p_{j+1,1}+a,...,p_{J1}+a,p_{J+1}+a,y+a\right)\text{ if }a\geq 0, \\ \sum_{k=1}^{j}q_k\left(p_{10}-a,...,p_{j0}-a,p_{j+1,1},...,p_{J1},p_{J+1},y\right)\text{ if }a<0. \end{array}\right. \end{aligned}$$

When $a \geq p_{J0} - p_{J1}$, then (33) reduces to

$$\begin{aligned}
& \Pr\left[\begin{array}{c} \max\left\{U_1\left(y-p_{11},\eta\right),...,U_J\left(y-p_{J1},\eta\right),U_{J+1}\left(y-p_{J+1},\eta\right)\right\} \\ \leq \max\left\{U_1\left(y+a-p_{10},\eta\right),...,U_J\left(y+a-p_{J0},\eta\right)\right\} \end{array}\right] \\
= \ & \Pr\left[\begin{array}{c} \max\left\{U_J\left(y-p_{J1},\eta\right),U_{J+1}\left(y-p_{J+1},\eta\right)\right\} \\ \leq \max\left\{U_1\left(y+a-p_{10},\eta\right),...,U_J\left(y+a-p_{J0},\eta\right)\right\} \end{array}\right] \\
= \ & 1-\Pr\left[\begin{array}{c} \max\left\{U_J\left(y-p_{J1},\eta\right),U_{J+1}\left(y-p_{J+1},\eta\right)\right\} \\ > \max\left\{U_1\left(y+a-p_{10},\eta\right),...,U_J\left(y+a-p_{J0},\eta\right)\right\} \end{array}\right] \\
\stackrel{(1)}{=} \ & 1-\Pr\left[\begin{array}{c} U_{J+1}\left(y-p_{J+1},\eta\right) \\ > \max\left\{U_1\left(y+a-p_{10},\eta\right),...,U_J\left(y+a-p_{J0},\eta\right)\right\} \end{array}\right] \\
= \ & \left\{\begin{array}{c} 1-q_{J+1}\left(p_{10},...,p_{J0},p_{J+1}+a,y+a\right) \text{ if } p_{J0}-p_{J1} \leq a,\ a \geq 0, \\ 1-q_{J+1}\left(p_{10}-a,...,p_{J0}-a,p_{J+1},y\right) \text{ if } p_{J0}-p_{J1} \leq a,\ a < 0, \\ 1, \text{ otherwise} \end{array}\right. .
\end{aligned}$$

Equality $\stackrel{(1)}{=}$ follows from the fact that if $a \geq p_{J0} - p_{J1}$, then $U_J\left(y - p_{J1}, \eta\right)$ must be no more than $U_J\left(y + a - p_{J0}, \eta\right)$. ■

The proof of EV is very similar and is omitted for the sake of brevity.