Nonparametric Approaches to Empirical Welfare Analysis

Debopam Bhattacharya*
University of Cambridge

November 2022

Abstract: Welfare analysis of policy interventions is ubiquitous in economic research. It plays an important role in merger analysis and antitrust litigation, design of tax and subsidy programs and informs the current debate on a universal basic income. In this paper, we provide a survey of existing empirical methods, based on cross-sectional micro-data, for calculating welfare-effects and deadweight loss resulting from realized or hypothetical policy change. We briefly outline classical parametric methods that are computationally tractable, and then discuss recently developed non-parametric approaches in greater detail. The latter avoid imposing statistical and functional-form restrictions on individual preferences. This makes the welfare estimates theoretically more credible, and clarifies exactly what welfare-relevant information is contained in demand distributions in various choice settings. However, these methods also demand greater in-sample variation in the data for practical implementation than classical parametric approaches. We then cover settings with externalities. The above results are theoretical, and take the demand function as known; therefore, we briefly discuss empirical problems around the estimation of demand itself and of welfare based on it. We conclude by suggesting areas for future research.

*I am grateful to Steven Durlauf, David Romer, Arthur Lewbel and three anonymous referees for feedback. Financial support from the European Research Council via Consolidator Grant EDWEL is gratefully acknowledged.

1 Introduction

Econometric program evaluation has become central to policy analysis in modern economics. The corresponding academic literature on identification of treatment effects (Heckman-Vytlacil 2007) and on optimal policy targeting (Manski 2004) has focussed on functionals of outcome distributions as the key object of interest for policy evaluation, while the problem of measuring program effects on unobservable *utilities* of individuals has been somewhat neglected in the econometric literature. For example, a school tuition subsidy program in a developing country would typically be evaluated in the econometric tradition via its "treatment effect" on aggregate enrolment, and the treatment choice problem would address the question of how to target a limited amount of subsidy funds to maximize this aggregate (Bhattacharya-Dupas 2012, Kitagawa-Tetenov 2018 and references therein). This approach neglects the question of how much and how differently do the individuals themselves value the subsidy – determined by their willingness to pay for school – and what is the subsidy's effect on aggregate utility, possibly weighted by the government's distributional preferences. In particular, consider a household that is both eligible for the subsidy, *and* would send their child to school irrespective of receiving any subsidy. Such a household would contribute nothing to the average treatment effect because receiving the subsidy will not change its schooling decision. But access to subsidized tuition would let them save the tuition expenditure and therefore increase their utility. These effects are missed if one focuses only on the effect of treatment on the outcome. Secondly, in many practical settings, multiple related outcomes of policy-interest are likely to be affected by a single intervention. For example, a price subsidy for mosquito-nets (Dupas 2014) can be evaluated in terms of aggregate take-up of nets, incidence of malaria, school absence of children and so forth. A welfare-based approach provides a natural way to aggregate these separate effects via how they determine the households' overall willingness-to-pay for the mosquito-net. Third, price-interventions for redistributions are often politically motivated. Efficiency-costs of such policies to society are therefore important metrics of political assessment.

Indeed, there is a large theoretical literature on social choice in welfare economics and a similarly rich body of theoretical research in public finance on optimal taxation and project evaluation where individual preferences play a central role. But attempts to bring these theories to real data, which inevitably involve dealing with unobserved heterogeneity in individual preferences, were relatively rare until recently.

The present survey provides an overview of nonparametric methods for empirical welfare analysis for a population with heterogeneous preferences. In contrast to classical parametric approaches, these methods avoid traditional functional-form assumptions, e.g. quasilinear utilities and income-effects, additive extreme-value distributed unobserved heterogeneity etc. It turns out that in many choice settings of practical importance, once the entire demand function is known or consistently estimated, fully nonparametric point-identification of welfare is possible without any restriction on preference heterogeneity; in other settings, one may only obtain bounds. From a theoretical angle, these results clarify what welfare relevant information is in fact contained in estimable demand functions under restrictions coming only from economic theory. A key finding from these results is that it is typically unnecessary to recover properties of *underlying preferences* from demand in order to conduct welfare analysis. The demand function itself is sufficient for that purpose. This insight pervades most of the results discussed here.[1] Note, however, that the welfare expressions are functionals of the underlying average demand, whose nonparametric identification/estimation requires greater variation of price and income in the data than traditional parametric approaches.

Our survey has the following structure. Each numbered line denotes a separate section. Sections 2-5 show how to obtain the expressions for welfare in terms of the structural demand function in different choice settings, i.e. how to calculate welfare if the underlying demand functions were known. Section 6 discusses identification of the demand function itself and practical data issues in estimating welfare from demand.

2 Discrete/Multinomial Choice

- 2.1 Parametric approach: McFadden's (1973) multinomial logit and Berry-Levinsohn-Pakes (1995)
- 2.2 Nonparametric approach
  - 2.2.1 Multinomial choice: welfare evaluation of
    - A. Price changes
    - B. Elimination/introduction of alternatives
    - C. Quality change
  - 2.2.2 Bundles, e.g. choice of package consisting of cable TV and/or broadband internet etc.
  - 2.2.3 Ordered choice, e.g. educational attainment (primary school, high school, college etc.)

3. Continuous choice

4. Empirical analysis of "social" welfare i.e. aggregate indirect utility, and optimal treatment allocation for maximizing it under budget constraints

- 4.1 Bergson-Samuelson social welfare under heterogeneity
- 4.2 Marginal welfare analysis
- 4.3 Social welfare in discrete choice settings

[1] This is reminiscent of Marschak (1953) who noted that "...for many problems of policy analysis, it is not necessary to identify fully specified structural models. All that is required to conduct many policy analyses... are policy invariant combinations of the structural parameters that are often much easier to identify than the individual parameters themselves and that do not require knowledge of individual structural parameters" (see Heckman 2011).

4.3.1 The Daly-Zachary-Williams theorem

4.3.2 Bhattacharya-Komarova (2022)

4.3.3 Binary choice

4.3.4 Optimal treatment allocation

4.3.5 Connection with Hicksian compensation

4.4 Ordered discrete choice and the continuous case

5. Welfare analysis under social interactions, e.g. welfare evaluation of health product subsidies when individual benefit from adoption depends on neighborhood adoption rate, causing targeted subsidies to have spillover effects

6. Identification and practical data issues in implementing welfare calculations with micro data, including (i) bounded support and shape restrictions, (ii) interval reporting of income, (iii) unobservability of price, and (iv) endogeneity of price/income

7. Future directions

   (a) Uncertainty and Insurance

   (b) Intra-household considerations

   (c) Nonlinear pricing

   (d) "Behavioral" welfare analysis

   (e) Using novel data sources

Two settings that are not included in the survey are (a) dynamic choice, which raises a host of unique definitional and behavioral questions with regard to welfare analysis, and also require the use of panel – as opposed to cross-sectional – data, and (b) contingent valuation methods based on the "stated preference" approach (Hanemann 1984), as opposed to the ''revealed preference" one, on which the present survey is based. In the "stated preference" approach, individuals' hypothetical willingness to pay for a proposed policy intervention is elicited via direct surveys, without data on actual behavior, as in the revealed preference methods. Economists have long been sceptical of this approach (Diamond-Hausman 1994), but it can be useful in two situations: (a) when realized policy variation is rare or nonexistent in the data and (b) when the goods under consideration are publicly rather than privately consumed, e.g. green spaces and recreation sites, for which typically no financial transaction data are usually available. We end this introduction by noting that a previous survey by Slesnick (1998) covers some of the social choice-theoretic issues and philosophical debates underlying common welfare measures and their aggregation, which are only briefly alluded to in the present survey.

2 Discrete choice

Formal econometric analysis of discrete choice, starting with McFadden (1974), incorporated welfare analysis as one of its goals. In the subsequent literature, welfare analysis was based on either restrictive assumptions on preference heterogeneity, e.g., quasilinear preferences implying absence of income-effects and equality of Hicksian and Marshallian welfare (Domencic-McFadden 1975, Small-Rosen 1981), or parametrically specified utility functions and heterogeneity distributions, (McFadden 1981, Herriges-Kling 1999, Petrin 2002, Dagsvik-Karlstrom 2005). A particularly widespread practice in applied work has been to use the so called logsum formula, (Small-Rosen 1981 and Train 2009, Anderson et al. 1992), which relies on strong parametric assumptions regarding both preferences and heterogeneity. Two key concerns with such parametric analyses are (i) model mis-specification leading to erroneous substantive conclusions, and (ii) identification of welfare distributions from functional form assumptions alone. Berry-Haile (2016, 2021), building on earlier results of Matzkin (2016), investigated non- and semiparametric identification of preference parameters in discrete choice models, but imposed quasilinear preferences when investigating welfare analysis. More recently, Bhattacharya (2015, 2018) derived some nonparametric results on empirical welfare analysis under general heterogeneity. These results show that in many – but not all – empirically relevant, discrete choice settings, welfare can be recovered nonparametrically from choice probabilities without restricting preference heterogeneity or functional form of (cardinal) utilities. In other cases, restrictions from the economic theory of utility maximization by budget-constrained consumers lead to bounds on average welfare and deadweight loss (cost of intervention less aggregate welfare change) In what follows, we provide a comprehensive exposition of the above points.

2.1 Parametric approach

We begin by describing some of the common parametric methods of demand and welfare analysis in multinomial choice models.

The classic McFadden model of multinomial choice consists of a setup with J mutually exclusive alternatives. The utility of consumer i from choosing alternative j is given by $U_{ij} = V(X_i, Z_{ji}; \theta) + \varepsilon_{ij}$ where X_i, Z_{ji} refer to observable characteristics of individual i and of alternative j faced by i respectively, $V(\cdot, \cdot, \cdot)$ is a function with known functional form indexed by an unknown parameter vector θ, and ε_{ij} is the unobservable (to the econometrician) component of utility with $(\varepsilon_{i1}, ..., \varepsilon_{iJ})$ being I.I.D. across i and following a Type 1 extreme value distribution. The dataset consists of n cross-sectional observations, where for each $i = 1, ..., n$, one observes the choice made by i among the J options and the characteristics of i and the characteristics z_{ji} for each $j = 1, ..., J$ faced by i. Typically, it is assumed that $V(X_i, Z_{ji}; \theta)$ is linear in x_i, z_{ji} with $-\alpha$ (an element of θ) denoting the coefficient of price of alternative j. One usually refers to α as the marginal utility of income which, like all the other coefficients, is assumed to be identical for all consumers. In this set up, the observable choice probabilities have the familiar multinomial logit form

$$\Pr\left[j|X_i = x, (Z_{1i}, ..., Z_{Ji}) = \left(z^1, ..., z^J\right)\right] = \frac{\exp\left(V\left(x, z^j; \theta\right)\right)}{\sum_{k=1}^{J} V\left(x, z^k; \theta\right)}$$

which gives rise to a likelihood function and θ can be estimated by MLE. The functional form assumptions above imply that

$$E\left[\max_j \left\{V\left(x_i, z^j; \theta\right) + \varepsilon_{ij}\right\}\right] = \frac{1}{\alpha}\ln\left(\sum_{j=1}^{J}\exp(V\left(x, z^j; \theta\right)\right) + C$$

where C is a constant. Now, if a policy intervention changes the characteristic vector from $\left\{z^{j0}\right\}$, $j = 1, ..., J^0$ to $\left\{z^{j1}\right\}$, $j = 1, ..., J^1$, then the resulting change in consumer surplus is given by

$$\frac{1}{\alpha}\left[\ln\left(\sum_{j=1}^{J^1}\exp(V\left(x, z^{j1}; \theta\right)\right) - \ln\left(\sum_{j=1}^{J^0}\exp(V\left(x, z^{j0}; \theta\right)\right)\right],$$

which is the well-known "logsum" formula derived in Small-Rosen (1981). The fact that J^0 need not equal J^1 reflects that the above expression covers the change in the number of alternatives as well. However, this formula depends on a large number of parametric assumptions, some of which are testable, e.g. the extreme value errors and the implied restrictive property of independence of irrelevant attributes (Hausman-McFadden 1984, Small-Hsiao 1985), additive errors (Bhattacharya 2021), quasilinear utilities and lack of income-effects, etc. Trajtenberg (1989) used the above specification to compute welfare effects of CT scanners in US hospitals. McFadden (1981) introduced more flexible parametric specification where coefficients of characteristics in $V(\cdot,\cdot,\cdot)$ were allowed to vary randomly across individuals.

Berry-Levinsohn-Pakes (1995) extended the McFadden model to a setting where only aggregate data on sale of J different varieties of a product are available from a cross-section of markets. In each market, one observes the characteristics of each variety including its price and its share among the total sale of all J varieties. A substantive extension offered in BLP is to allow for unobservable product characteristics and demand shocks which can vary across markets. The latter are likely to make prices endogenous, and BLP developed methods to instrument for this and obtain consistent estimates of the preference parameters using the assumed extreme value distribution of additive idiosyncratic unobservables.

Based on the BLP method, Petrin (2002) measured the welfare effects of introduction of minivans in the US automobile market via the compensating variation in a fully parametric model and using micro data on consumers. He retained the extreme value distributed errors and allowed for income-effects by replacing the term $-\alpha p_j$ in $V(x_i, z_j; \theta)$ to $-\alpha_i \ln(y_i - p_j)$ where y_i refers to individual i's income and α_i, which measures how utility depends on the numeraire, is allowed to take three distinct values depending on y_i. He calculated the expected compensating variation under these assumptions, and found them to be much smaller and more robust to parametric distributional assumptions than those obtained using only aggregate market data. Herriges-Kling (1999), on the other hand, provided computationally intensive bounds on welfare estimates of price changes in the recreational fishing industry in California by generalizing the canonical extreme value errors to the Generalized Extreme Value (GEV) distribution and allowing for income-effects using a similar expression as Petrin's. Both these studies were fundamentally parametric and the identification of the relevant welfare distribution followed from identification of the underlying structural parameters via maximum likelihood in fully parametric models.

2.2 Nonparametric approach

We now move to the main focus of this survey namely the nonparametric approach to welfare analysis covering the more recent developments in this literature.

2.2.1 Multinomial choice

Consider a standard multinomial choice setting with J alternatives indexed by j, the price vector $\mathbf{p}$ and income y. Economic applications usually require welfare analysis of three different types of economic changes in this setting, namely (a) price changes, (b) addition or elimination of alternatives to the individuals' choice sets, and (c) quality change.

Let $U_j(y-p_j,\eta)$ denote the utility from buying alternative j at price p_j, with η denoting (unrestricted) unobserved preference heterogeneity. 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

$$\begin{aligned} q_j(\mathbf{p},y) &= \Pr{}_\eta \left[U_j(y-p_j,\eta) > \max_{k\neq j}\{U_k(y-p_k,\eta)\}\right] \\ &\overset{defn.}{=} \int 1\left\{U_j(y-p_j,r) > \max_{k\neq j}\{U_k(y-p_k,r)\}\right\} dF_\eta(r), \end{aligned} \tag{1}$$

where $\Pr_\eta$ refers to probability calculated with respect to the marginal distribution of η. 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. Throughout this section, we will maintain the following assumption which states that preferences are non-satiated with respect to the numeraire. Note that this assumption leaves the dimension of heterogeneity completely unspecified, and says nothing about how utility changes with unobserved heterogeneity.

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

We now state expressions for welfare effects of three common economic changes in the above setting, namely, price changes, elimination/addition of alternatives and change in product attributes. These expressions characterize the distribution of welfare, or bounds on it, in terms of the choice probability function $q_j(\mathbf{p},y)$. Therefore, *once the $q_j(\cdot,\cdot)$'s are identified on the support of price* *and income* (discussions on identification of $q_j(\cdot,\cdot)$ is postponed to Section 6 below), so are the welfare distributions.

A. Price change Consider a price change from $\mathbf{p}_0 \equiv (p_{10}, p_{20}, ..., p_{J0})$ to $\mathbf{p}_1 \equiv (p_{11}, p_{21}, ..., p_{J1})$. Assume without loss of generality that

$$p_{J1} - p_{J0} \geq p_{J-1,1} - p_{J-1,0} \geq ... \geq p_{11} - p_{10}. \tag{2}$$

Then the equivalent variation at income y for an η type consumer is the income reduction S in the initial situation that would lead to attainment of the indirect utility achieved at the final prices. Formally, the Hicks-Kaldor (Hicks 1943, Kaldor 1939) equivalent variation is the solution S to the equation:

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

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

$$\begin{aligned} & \max\left\{U_1\left(y+S-p_{11},\eta\right), U_2\left(y+S-p_{21},\eta\right), ..., U_J\left(y+S-p_{J1},\eta\right)\right\} \\ = \; & \max\left\{U_1\left(y-p_{10},\eta\right), U_2\left(y-p_{20},\eta\right), ..., U_J\left(y-p_{J0},\eta\right)\right\}. \end{aligned} \tag{4}$$

The value of S solving (3) or (4) can take negative values if one or more of the final prices are lower than their initial counterparts.

As η varies in the population, the compensating variation and equivalent variation will have a distribution across consumers. The following theorem from Bhattacharya (2018) establishes the CDF of the compensating variation in terms of the structural choice probability function.[2]

Theorem 1 *Consider the multinomial choice setup 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 (2). Denote $p_{j1} - p_{j0}$ by Δp_j for $j = 1, ..., J$. Under assumption 1, the marginal distribution of the individual equivalent*

[2] It is implicit here that the price changes do not alter the distribution of heterogeneity, 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, as is implicitly assumed in this literature.

variation evaluated at income y is given by

$$
\begin{aligned}
&\Pr(EV \leq a) \\
= &\begin{cases}
0 \textit{ if } a < \Delta p_1, \\
\sum_{k=1}^{j} q_k \begin{pmatrix} p_{11}, ..., p_{j1}, \\ p_{j+1,0} + a, ..., p_{J0} + a, y \end{pmatrix} \textit{ if } \Delta p_j \leq a < \Delta p_{j+1},\ a \geq 0,\ 1 \leq j \leq J-1, \\
\sum_{k=1}^{j} q_k \begin{pmatrix} p_{11} - a, ..., p_{j1} - a, \\ p_{j+1,0}, ..., p_{J0}, y - a \end{pmatrix} \textit{ if } \Delta p_j \leq a < \Delta p_{j+1},\ a < 0,\ 1 \leq j \leq J-1, \\
1 \textit{ if } a \geq \Delta p_J,
\end{cases}
\end{aligned}
\tag{5}
$$

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

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

where $q_k s$ are defined above in equation (1). (The separate entries for $a \geq 0$ and $a < 0$ in each line of (5) and (6) arise from accommodating rise and fall of prices, respectively).

The proof of this result starts with the observation that the RHS of (4) is strictly increasing and continuous in S, by Assumption 1. Therefore, the event $S \leq a$ is equivalent to the event

$$
\begin{aligned}
& \max\{U_1(y - p_{10}, \eta), U_2(y - p_{20}, \eta), ..., U_J(y - p_{J0}, \eta)\} \\
\leq\ & \max\{U_1(y + a - p_{11}, \eta), U_2(y + a - p_{21}, \eta), ..., U_J(y + a - p_{J1}, \eta)\}.
\end{aligned}
\tag{7}
$$

Now, depending on where a lies within the ordered set $\{\Delta p_J \geq \Delta p_{J-1} \geq ... \geq \Delta p_1\}$, we get the different cases that appear in (6). For example, if $\Delta p_1 \leq a < \Delta p_2$, then (7) reduces to

$$
\begin{aligned}
& \max\{U_2(y - p_{20}, \eta), ..., U_J(y - p_{J0}, \eta)\} \leq U_1(y + a - p_{11}, \eta) \\
\Leftrightarrow\ & \max\{U_2(y + a - (p_{20} + a), \eta), ..., U_J(y + a - (p_{J0} + a), \eta)\} \leq U_1(y + a - p_{11}, \eta)
\end{aligned}
$$

whose probability equals $q_1(p_{11}, p_{20} + a, ...p_{J0} + a, y + a)$.

The above 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 (5) becomes

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

The expected welfare change and deadweight loss follow directly from the CDF. For example, if (8) results from a subsidy $|\Delta p_1|$ on alternative 1, and a tax Δp_3 on alternative 3, then

$$E(Deadwt\_Loss) = E\left(EV\right) - \Delta p_1 \times q_1\left(p_{11}, 0, p_{31}, y\right) - \Delta p_3 \times q_3\left(p_{11}, 0, p_{31}, y + \Delta p_3\right),$$

where $E\left(EV\right)$ can be calculated directly from (8) using the well-known result that for a random variable X with CDF $F\left(\cdot\right)$ and finite mean,

$$E\left(X\right) = \int_0^{\infty} \left(1 - F\left(x\right)\right) dx - \int_{-\infty}^{0} F\left(x\right) dx.$$

The expressions derived above have been used in several recent empirical studies, e.g. Hninn et al. 2017 to evaluate a water improvement policy in Myanmar, Kroft et al. (2020) who compute the rents arising from imperfect competition in the US construction industry and Kamat-Norris (2021), who calculate welfare effects of school vouchers in Washington DC.

B. 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 (i.e. compensating and equivalent variation) that would result from eliminating the $J+1^{st}$ alternative. Applied researchers typically use the idea that eliminating an alternative is like raising its price to infinity, and thereafter use the welfare formulae for price change to evaluate welfare effects of eliminating an alternative. Our analysis below provides a formal justification for this approach, and shows that it is valid only when prices of other alternatives do not change and when one is interested in the equivalent variation.

Toward that end, consider the same set up as the multinomial case above with $J+1$ alternatives. 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 $(p_{10}, ..., p_{J0}, p_{J+1})$, following elimination of the $J+1^{st}$ alternative, the eventual price vector becomes $(p_{11}, ..., p_{J1})$. The marginal distributions of equivalent variation and compensating variation resulting from this change are given by the following corollary.

Corollary 1 *Suppose Assumption 1 holds for all alternatives. Also assume that for any bundle* $(j, y - p_j)$, *if* $p_j \uparrow \infty$ *with prices of other alternatives finite, then for each* η *at least one other bundle* $(k, y - p_k)$ *will be strictly preferred to* $(j, y - p_j)$. *Let* $q_k\left(\cdot, ..., \cdot, y\right)$ *be as defined in (1), with*

$J+1$ alternatives. Then

$$\Pr(CV \leq a) = \begin{cases} 0 \text{ } if \text{ } a < \Delta p_1, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{11}, ..., p_{j1}, \\ p_{j+1,0} + a, ..., p_{J0} + a, p_{J+1} + a, \\ y + a \end{pmatrix}, \text{ } if \text{ } \Delta p_j \leq a < \Delta p_{j+1}, \text{ } j = 1, .., J-1, \text{ } a \geq 0, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{11} - a, ..., p_{j1} - a, \\ p_{j+1,0}, ..., p_{J0}, p_{J+1}, y \end{pmatrix}, \text{ } if \text{ } \Delta p_j \leq a < \Delta p_{j+1}, \text{ } 1 \leq j < J-1, \text{ } a < 0, \\ 1 - q_{J+1}(p_{11}, ..., p_{J1}, p_{J+1} + a, y + a), \text{ } if \text{ } \Delta p_J \leq a, \text{ } a \geq 0, \\ 1 - q_{J+1}(p_{11} - a, ..., p_{J1} - a, p_{J+1}, y), \text{ } if \text{ } \Delta p_J \leq a < 0. \end{cases} \tag{9}$$

On the other hand,

$$\Pr(EV \leq a) = \begin{cases} 0 \text{ } if \text{ } a < p_{11} - p_{10}, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{11}, ..., p_{j1}, \\ p_{j+1,0} + a, ..., p_{J0} + a, p_{J+1} + a, y \end{pmatrix}, \text{ } if \text{ } \Delta p_j \leq a < \Delta p_{j+1}, \text{ } j = 1, .., J-1, \text{ } a \geq 0, \\ \sum_{k=1}^{j} q_k \begin{pmatrix} p_{11} - a, ..., p_{j1} - a, \\ p_{j+1,0}, ..., p_{J0}, p_{J+1}, y - a \end{pmatrix}, \text{ } if \text{ } \Delta p_j \leq a < \Delta p_{j+1}, \text{ } 1 \leq j < J-1, \text{ } a < 0, \\ 1 - q_{J+1}(p_{11}, ..., p_{J1}, p_{J+1} + a, y), \text{ } if \text{ } \Delta p_J \leq a, \text{ } a \geq 0, \\ 1 - q_{J+1}(p_{11} - a, ..., p_{J1} - a, p_{J+1}, y - a), \text{ } if \text{ } \Delta p_J \leq a < 0. \end{cases} \tag{10}$$

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 $J+1^{st}$ 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. This can be viewed as the additional cost of nonparametric methods relative to tractable and parsimonious parametric approaches where identification of (finite-dimensional) parameters typically requires only minimal price/income variation.

Corollary 2 *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*

$$\Pr(CV \leq a) = \begin{cases} 0 \text{ } if \text{ } a < 0, \\ 1 - q_{J+1}(p_{10}, ..., p_{J0}, p_{J+1} + a, y + a) \text{ } \text{ } if \text{ } 0 \leq a. \end{cases} \tag{11}$$

$$\Pr(EV \leq a) = \begin{cases} 0 \text{ } if \text{ } a < 0, \\ 1 - q_{J+1}(p_{10}, ..., p_{J0}, p_{J+1} + a, y) \text{ } \text{ } if \text{ } 0 \leq a. \end{cases} \tag{12}$$

The corresponding average values are given by

$$E\left(CV\right) = \int_{p_{J+1}}^{\infty} q_{J+1}\left(p_{10},...,p_{J0},r,y+r-p_{J+1}\right)dr, \tag{13}$$

$$E(EV) = \int_{p_{J+1}}^{\infty} q_{J+1}\left(p_{10},...,p_{J0},r,y\right)dr, \tag{14}$$

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

The expression (14) is commonly used as a 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 $E\left(CV\right)$ 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 (5) 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 (14) are identified directly from functional form assumptions. While this may be viewed as a practical advantage of parametric approaches, it is not clear that these parametrically obtained point-estimates (as opposed to nonparametrically obtained bounds) are reliable, due to the theoretically unjustified assumptions underlying their specification, e.g. extreme value distributed unobservables.

New good: Note that the above expressions can be used for retrospective calculation of welfare distributions corresponding to *introduction* of a new alternative, simply by interchanging the labels of equivalent variation and CV. Here "retrospective" means that we have consumption data from both before and after the product has been introduced. In particular, suppose that starting with the set of options $\{1,...,J\}$, the alternative $J+1$ is introduced subsequently, which changes prices from $(p_{11},...,p_{J1})$ to $(p_{10},...,p_{J0},p_{J+1})$. Suppose that we have data on a cross-section of individual choices in the pre-introduction and from another cross-section in the post-introduction situation. Then the distribution of the equivalent variation and compensating variation resulting from introduction of alternative $J+1$ are given by (9) and (10), respectively.

C. Quality change We now consider welfare calculation when characteristics of a choice alternative changes. For example, in transportation choice settings, increasing the frequency of buses would raise every individual's utility from choosing the bus option, as opposed to driving or walking. These characteristics, unlike prices, do not appear as additive components of net income inside utilities. The additivity of price, resulting from the budget constraint, was instrumental in deriving the results of the previous section. Therefore, welfare analysis of changes in characteristics would require a different result.

To keep the exposition transparent, we will focus on binary choice, generalization to the multinomial case is straightforward. Accordingly, suppose an individual at income y and preference η is choosing between two options 0 and 1, with utilities given by $U_0(y,\eta)$ and $U_1(y-p,x,\eta)$, where x represents a vector of characteristics of option 1. For example, in the transportation example, option 0 is walking, option 1 is taking the bus, p is the bus fare, and x is the frequency of the bus service. Define the structural choice probability of choosing the bus at price p, income y and frequency x as

$$q_1(p,y,x)=\int 1\{U_1(y-p,x,r)>U_0(y,r)\}\,dF_\eta(r), \tag{15}$$

where $F_\eta(\cdot)$ represents the marginal distribution of unobserved heterogeneity. Analogous to assumption 1 above, suppose that

Assumption 2 *For each x and η, $U_1(\cdot,x,\eta)$ is strictly increasing and continuous; additionally, for each η, $U_0(\cdot,\eta)$ is strictly increasing and continuous.*

Now, suppose from an initial price p_0 and initial quality x_0, the price and attribute change to p_1 and x_1, respectively. The individual compensating variation resulting from this change is defined by the solution S to the equation

$$\max\{U_0(y,\eta),U_1(y-p_0,x_0,\eta)\}=\max\{U_0(y+S,\eta),U_1(y+S-p_1,x_1,\eta)\}. \tag{16}$$

It turns out that the distribution of S cannot be non-parametrically identified from knowledge of the structural choice probabilities. Bhattacharya (2018) establishes this via a counter-example where two different distributions of η produce identical choice probabilities at each (p,y) but the distribution of the compensating variation is different in these two cases.

One can, however, construct bounds on these distributions. To see where these bounds come from, note from (16) and Assumption 2, that for $a<0$

$$\begin{aligned}
&\Pr(CV\le a)\\
=\;&\Pr\left[\begin{array}{c}\max\{U_0(y,\eta),U_1(y-p_0,x_0,\eta)\}\\ \le \max\left\{U_0\left(\underbrace{y+a}_{<y},\eta\right),U_1(y+a-p_1,x_1,\eta)\right\}\end{array}\right]\\
=\;&\Pr[\max\{U_0(y,\eta),U_1(y-p_0,x_0,\eta)\}\le U_1(y+a-p_1,x_1,\eta)].
\end{aligned}$$

The Frechet upper bound on this is given by

$$\begin{aligned}
&\min\left\{\begin{array}{l}\Pr[U_0(y,\eta)\le U_1(y+a-p_1,x_1,\eta)],\\ \Pr[U_1(y-p_0,x_0,\eta)\le U_1(y+a-p_1,x_1,\eta)]\end{array}\right\}\\
=\;&\min\left\{\begin{array}{l}q_1(p_1-a,y,x_1),\\ \Pr[U_1(y-p_0,x_0,\eta)\le U_1(y+a-p_1,x_1,\eta)]\end{array}\right\}.
\end{aligned}$$

The ultimate upper bound comes from a lower bound on $\Pr\left[U_1\left(y-p_0,x_0,\eta\right)>U_1\left(y+a-p_1,x_1,\eta\right)\right]$, obtained using

$$\begin{aligned}&\Pr\left[U_1\left(y-p_0,x_0,\eta\right)>U_1\left(y+a-p_1,x_1,\eta\right)\right]\\ \geq\ &\max_{y'}\Pr\left[U_1\left(y-p_0,x_0,\eta\right)\geq U_0\left(y',\eta\right)\geq U_1\left(y+a-p_1,x_1,\eta\right)\right],\end{aligned}$$

and then applying Frechet bounds to this.

Weak Separability and point-identification: It turns out that a restriction on preferences, far short of fully parametric specifications, can yield point-identification of welfare distributions for quality change. This is where the utility from alternative 1 is weakly separable in unobserved heterogeneity, i.e., $U_1\left(y-p,x,\eta\right)\equiv V_1\left(h_1\left(y-p,x\right),\eta\right)$, where $V_1\left(\cdot,\eta\right)$ is strictly increasing for each η, and $h_1\left(\cdot,\cdot\right)$ is continuous and strictly increasing in the first component. A special case of weak separability, where the functional form of $h_1\left(\cdot,\cdot\right)$ is assumed *known* up to estimable real parameters is the familiar "index restriction", widespread in the semiparametrics literature. Weak separability has been used previously in econometrics for identifying effects of endogenous dummy variables (Vytlacil-Yildiz 2008) and nonlinear hedonic models (Heckman et al. 2010, Theorem 3.1), while index restrictions have been used for achieving identification in general demand systems, (Berry-Haile 2015).

Now consider a change from (p_0,x_0) to (p_1,x_1) satisfying the following condition:

Condition S: *Assume that the change from (p_0,x_0) to (p_1,x_1) is small enough that there exists a finite a satisfying*

$$h_1\left(y+a-p_1,x_1\right)=h_1\left(y-p_0,x_0\right).$$

This basically says that the quality and price change are not so large that no amount of compensation can restore the utility from consuming alternative 1 to the pre-change level for any η.

Proposition 1 *Suppose weak separability holds. Consider a change from (p_0,x_0) to (p_1,x_1) satisfying condition (S). Then there is a unique $\bar{a}$ that solves $q_1\left(p_1-\bar{a},y,x_1\right)=q_1\left(p_0,y,x_0\right)$. Furthermore, the distribution of individual compensating variation resulting from this change is point-identified from the structural choice probabilities, and it is given by the following expressions:*

If $\bar{a}\leq 0$, then

$$\Pr\left(S\leq a\right)=\begin{cases}0,\ \text{if } a<\bar{a},\\ q_1\left(p_1-a,y,x_1\right),\ \text{if } \bar{a}\leq a<0,\\ 1,\ \text{if } a\geq 0.\end{cases}$$

If $\bar{a}>0$, then

$$\Pr\left(S\leq a\right)=\begin{cases}0,\ \text{if } a<0,\\ q_0\left(p_0+a,y+a,x_0\right),\ \text{if } 0\leq a<\bar{a},\\ 1,\ \text{if } a\geq \bar{a}.\end{cases}$$

See Bhattacharya (2018) for a proof.

2.2.2 Bundles and Complements

In many practical settings consumers have to choose among bundles of non-exclusive options (Hendel 1999). For example, suppose option 1 is a sports package offered by a cable TV network which costs P_1 and choice 2 is a movie package which costs P_2. A household with income Y and preferences η then faces a choice among four exclusive bundles – $\{1\},\{2\},\{1,2\},\{0\}$ (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. In this setting, Bhattacharya (2018) shows that for a change in price of one alternative, welfare distributions are point-identified, but for changes in both prices, they are not.

To see why, consider the compensating variation 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 compensating variation evaluated at income $Y=y$ is the solution to the equation

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

Now, group option $\{1\}$ and $\{1,2\}$ together (call it group A) and options $\{0\}$ and $\{2\}$ together and call it group B. Define

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

Then (17) becomes

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

If the U functions are strictly increasing in the first argument for each η, then so are $V_A(\cdot,\varepsilon)$ and $V_B(\cdot,\varepsilon)$ for each ε. Now we can apply Theorem 1 above to get the marginal distribution of the compensating variation, which is given by

$$\Pr(CV\le a) = \begin{cases} 0 \text{ if } a<0, \\ q_0(p_{10}+a,p_2,y+a)+q_2(p_{10}+a,p_2,y+a), \text{ if } 0\le a<p_{11}-p_{10}, \\ 1 \text{ if } a\ge p_{11}-p_{10}. \end{cases} \tag{19}$$

However, if both P_1 and P_2 change in the above setting, the marginal distribution of the resulting compensating variation is no longer point-identified. This was shown by Bhattacharya (2018) through a counter-example similar to the quality change case above, where two different joint distribution of random coefficients produce identical choice probabilities at each price and income but different expressions for the compensating variation distribution.

2.2.3 Ordered Choice

The multinomial choice scenario described above may be contrasted with a situation of ordered choice, i.e., where a good can be bought in discrete units of 0, 1, 2, etc. *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 non-zero 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-Newey (2016), can be viewed as a limiting case of ordered choice with uniform unit price. To see the identification failure, let $U_0(y, \eta)$, $U_1(y-p, \eta)$ and $U_2(y-2p, \eta)$ denote the utility from buying 0, 1 or 2 units of a good respectively, where p denotes the (uniform) unit price and the household's income is y. A second feature that distinguishes ordered choice from the multinomial case is that under non-satiation in the discrete good, we must have for any $a > 0$, $\Pr[U_1(a, \eta) < U_2(a, \eta)] = 1$, i.e. if one had the same amount a of numeraire in both cases, 2 units of the discrete good (e.g. 2 apples) must give higher utility than 1 for every consumer. There is no such requirement for multinomial choice, i.e. there is no restriction on $\Pr[U_j(\cdot, \eta) \leq U_{j+1}(\cdot, \eta)]$ in (1), where the labels j are arbitrary

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

$$\begin{aligned} & \max\{U_0(y, \eta), U_1(y-p_0, \eta), U_2(y-2p_0, \eta)\} \\ = & \max\left\{U_0\left(y+S^{CV}, \eta\right), U_1\left(y+S^{CV}-p_1, \eta\right), U_2\left(y+S^{CV}-2p_1, \eta\right)\right\}. \end{aligned} \tag{20}$$

Consider the probability that $S^{CV} = p_1 - p_0$. If the utility differences are continuously distributed, then the only situation where $S^{CV} = p_1 - p_0$ is where the maximum on the LHS of (20) is $U_1(y-p_0, \eta)$ and that on the RHS is $U_1\left(y+S^{CV}-p_1, \eta\right)$.[3] Thus

$$\begin{aligned} & \Pr\left[S^{CV} = p_1 - p_0\right] \\ = & \Pr\left[\begin{array}{l} U_1(y-p_0, \eta) \geq \max\{U_0(y, \eta), U_2(y-2p_0, \eta)\}, \\ U_1(y+p_1-p_0-p_1, \eta) \geq \max\{U_0(y+p_1-p_0, \eta), U_2(y+p_1-p_0-2p_1, \eta)\} \end{array}\right] \\ = & \Pr[U_1(y-p_0, \eta) \geq \max\{U_0(y+p_1-p_0, \eta), U_2(y-2p_0, \eta)\}], \end{aligned} \tag{21}$$

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 (21) is positive. However,

[3] For example, if instead the first term is the maximum for the LHS of (20) and the second term is the maximum for the RHS of (20) with $S = p_1 - p_0$, then

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

which will have zero probability if $U_1(y-p_0, \eta) - U_0(y, \eta)$ is continuously distributed, e.g., 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.

it is clear from (21) that the probability that $S^{CV} = p_1 - p_0$ 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 no (p^*, y^*) which satisfies all three equations simultaneously. So the distribution of the compensating variation cannot be nonparametrically point-identified.

An important lesson to emerge from the various cases discussed above is that parametric approaches to welfare analysis cited above do not clarify whether the identification of welfare 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. For example, the fact that nonparametric point-identification results, which hold for *unordered* multinomial choice, fail for *ordered* choice would not be apparent if one were to focus only on parametric models. Gentzkow (2007) calculated welfare effects of introducing an alternative with possible complementarity with an existing option in a discrete choice model without income-effects and additive extreme value distributed errors. Given the discussion above, it is apparent that average welfare is not identified in that setting without these parametric restrictions.

3 Continuous choice

Empirical welfare analysis of price change for a continuous good, such as raising a gasoline tax (Poterba 1991) or removing a rice subsidy (Jha and Srinivasan 2001), began with the seminal papers of Hausman (1981) and Vartia (1983). An up-to-date summary of this literature appears in Hausman-Newey (2017). Welfare analysis in this case is based on Roy's identity which is a partial differential equation linking the unobservable indirect utility functions $V(p, y)$ (the basis of welfare calculations) with observable demand functions $q(p, y)$, namely

$$-\frac{\partial V(p, y)}{\partial p} / \frac{\partial V(p, y)}{\partial y} = q(p, y),$$

with p, y denoting unit price and individual income, respectively. The RHS of the above equation is estimable (nonparametrically) from the data, whence the PDE is solved for $V(\cdot, \cdot)$, which then leads to the compensating and equivalent variation resulting from a price change from p_0 to p_1 at income y.

If unobserved preference heterogeneity (denoted by η) is allowed, however, demand analysis in general becomes more subtle (Lewbel 2001). In particular, Roy's identity for an individual type η consumer take the form

$$-\frac{\partial V(p,y,\eta)}{\partial p} / \frac{\partial V(p,y,\eta)}{\partial y} = q(p,y,\eta), \tag{22}$$

whence average welfare can be calculated using the Hausman-Vartia type approach under only very specific preference structures, such as the Gorman polar form. But if preference heterogeneity is left unrestricted, then Hausman-Newey (2016) have shown that the distribution of individual demand does not point-identify the average welfare effect of a price change. The basic argument is that cross-sectional demand data cannot distinguish between two scenarios, namely (A) where for each $\tau \in [0,1]$, the τth quantile of demand at each (p,y) represents the demands of consumer τ whose rank in the demand distribution remains unchanged when (p,y) change, and (B) when the rank of any consumer in the demand distribution can change with changes in (p,y), so that the observable τth quantile of demand cannot be associated with a specific consumer across all (p,y). For any specific price change and at a fixed income, the average equivalent variation is generically different in cases (A) and (B), but they are not empirically distinguishable. Hausman-Newey (2016) show that (sharp) bounds may be constructed on average welfare by assuming prior bounds on the magnitude of income-effects. The heuristic way to see where these bounds come from is as follows. Note that by definition of the compensating variation $S(p,y,\eta)$, the indirect utility $V(p, y+S(p,y,\eta), \eta)$ remains constant as p varies. Differentiating with respect to p and using (22) yields the PDE

$$\frac{\partial S(p,y,\eta)}{dp} = -q(p, y+S(p,y,\eta), \eta), \ S(p_0,y,\eta) = 0.$$

Then expand the RHS around y as

$$\frac{\partial S(p,y,\eta)}{dp} = -q(p,y,\eta) - S(p,y,\eta) \times \frac{\partial}{\partial y} q\left(p, y+\tilde{S}(p,y,\eta), \eta\right), \ S(p_0,y,\eta) = 0.$$

A uniform bound on $\frac{\partial}{\partial y} q\left(p, y+\tilde{S}(p,y,\eta), \eta\right)$ will then produce bounds on the compensating variation. Evidently, the width of the bounds is increasing in the size of the income-effect. Hausman-Newey (2016) also show that these bounds are less sensitive to income effect bounds when a smaller proportion of income is spent on the good. In an application to gasoline purchase in the United States, Hausman-Newey 2016 find income-effects to be very small, which yields tight bounds on the nonparametrically estimated average equivalent variation and deadweight loss resulting from a hypothetical tax induced price change.

4 Social welfare

The Hicksian measures discussed above are often used in empirical IO, trade and healthcare applications to evaluate actual or potential interventions. Examples include Petrin (2002) who evaluates introduction of minivans in the US car market, Goolsbee-Petrin (2004) who examine availability of satellite TV in addition to cable TV, Porto (2006) who assesses the Mercosur trade reforms in Argentina, De Nardi et al. (2016) who evaluate access to Medicaid insurance for retirees, and so forth. While Hicksian measures are useful ways to measure change in *individual* utility and its distribution in the population, adding them across consumers to obtain a measure of *social* welfare has several undesirable features. These include (i) an implicit assumption of a constant social marginal utility of income across income levels, i.e. that an additional dollar is valued by society in the same way no matter whether a rich or a poor person gets it,[4] (Blackorby-Donaldson 1988, Dreze 1998, Banks-Blundell-Lewbel 1996), (ii) comparing allocations via aggregate compensation suffers from conceptual ambiguities like Scitovszky reversal of social preferences (Mas-Colell et al. 1995, page 830-31), as illustrated in the following figure[5]

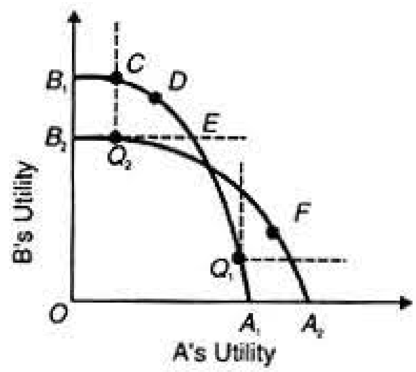


Fig. 2

The figure shows two allocations Q_1 and Q_2 with the utility possibility frontiers $A_1Q_1DB_1$ and $A_2FQ_2B_2$ through them intersecting. Each frontier represents the utility combinations attainable via redistribution between individuals A and B, starting from any point on it. Then the allocation D Pareto dominates Q_2 and can be attained from Q_1 via redistribution. Therefore Q_1 is *superior*

[4] This can be seen as follows. Social welfare $V(p) = \int H(W(p,y,\eta))\,dF(y,\eta)$, where $H(\cdot)$ denotes the welfare weight function of the planner, implying

$$\begin{aligned} V'(p) &= H'(W(p,y,\eta))\frac{\partial W(p,y,\eta)}{\partial p} \\ &\overset{\text{Roy's identity}}{=} -q(p,y,\eta)\times\underbrace{\left[H'(W(p,y,\eta))\times\frac{\partial W(p,y,\eta)}{\partial y}\right]}_{\text{marginal social utility of income}}, \end{aligned}$$

where $q(p,y,\eta)$ is ordinary demand at price p and income y, so that

$$V(p_0,\eta) - V(p_1,\eta) = \int_{p^0}^{p^1}[MSU(p,y,\eta)\times q(p,y,\eta)]\,dp,$$

whereas the average compensating variation equals $\int_{p^0}^{p^1} q^c(p, W(p_0,y,\eta),\eta)\,dp$, where q^c denotes compensated demand. It is only when MSUI equals 1 and compensated and Marshallian demand coincide, that these two expressions are equals.

[5] This graph is taken from https://www.economicsdiscussion.net/welfare-economics/criterion-of-welfare-with-diagram/18937

to Q_2 via the aggregate compensation principle. At the same time, the allocation F which can be attained via redistribution from Q_2 is Pareto superior to Q_1, implying that Q_1 is *inferior* to Q_2 via the compensation principle; so aggregate CV is again negative, thus leading to an ambiguity. These conceptual shortcomings of the Hicksian approach have led instead to widespread use of the Bergson-Samuelson average social welfare criterion in applied research in public finance. In that literature, cost-benefit analysis of an intervention was traditionally conducted by comparing the expenditure on it with the change it brings about in aggregate indirect utility (Bergson 1954, Samuelson 1947 and Mirrlees 1971). However, when bringing these concepts to data, this classical literature ignored unobserved heterogeneity and imposed functional form restrictions on the utility functions of 'representative' consumers who were assumed to vary only in terms of observables, see Deaton (1984) and Ahmad-Stern (1984). Later work such as Feldstein (1999), Saez (2001) have shown that in labour supply models with consumption leisure trade-off and heterogeneous agents, the optimal income-tax rate depends on individual heterogeneity via certain aggregates only, such as the average elasticity of taxable income with respect to the marginal tax rate, where the taxable income distribution is endogenously determined with the tax rate. For empirical implementation, these results require parametric modelling of preferences, such as Saez (2001) page 219 and Section 5. Similarly, Manski (2014) derives bounds on optimal income-tax schedule when consumers have heterogeneous Cobb-Douglas preferences. In an econometric sense, these are not nonparametric 'identification' results that express the object of interest (consumer welfare, optimal tax schedule etc.) in terms of quantities directly estimable from the data without making functional form assumptions about unobservables. The rest of this section discusses relatively recent methods that avoid making such assumptions.

4.1 Average welfare under nonparametric heterogeneity

The Bergson-Samuelson approach (see Bergson 1954, Atkinson 1970) evaluates alternative price-income situations via an aggregate indirect utility criterion. In particular, suppose $W(\mathbf{p}, y, \eta)$ denotes the indirect utility achieved at price $\mathbf{p}$ for individuals with income y and unobserved heterogeneity η. Then the average (over unobserved heterogeneity) social welfare at income y and prices $\mathbf{p}$ is given by

$$\bar{W}^{\varepsilon}(\mathbf{p}, y) \equiv \int \frac{W(\mathbf{p}, y, \eta)^{1-\varepsilon}}{1-\varepsilon} dF(\eta) = E_{\eta}\left\{\frac{W(\mathbf{p}, y, \eta)^{1-\varepsilon}}{1-\varepsilon}\right\}, \tag{23}$$

where E_{η} denotes expectation taken with respect to the marginal distribution of η, and $\varepsilon \in [0, 1)$ captures the planner's redistributional preferences, with low values of ε denoting stronger preference for redistribution from the rich to the poor. Accordingly, two price income joint distributions F and G can be compared in terms of average social utility via their mean welfare, i.e. $\int \bar{W}^{\varepsilon}(\mathbf{p}, y)\, dF(\mathbf{p}, y)$ and $\int \bar{W}^{\varepsilon}(\mathbf{p}, y)\, dG(\mathbf{p}, y)$.

4.2 Marginal welfare analysis

Marginal welfare analysis is an alternative to the average welfare approach. It attempts to approximate, nonparametrically, effects of *small* interventions on average social welfare (Ahmad-Stern 1984, Mayshar 1990, Finkelstein-Hendren 2020). The pivotal concept is that of the 'marginal value of public funds' (MVPF), defined as the ratio of beneficiaries' marginal willingness to pay for a policy change to the marginal cost of the intervention to the government. This approach does not cover non-marginal interventions and does not clarify how to account for income-effects and unobserved preference heterogeneity across individuals targeted by such non-marginal interventions. Further, the larger the behavioral impact of the intervention, the poorer of course is the resulting approximation through the marginal approach.

In what follows, we outline the average welfare approach first for multinomial and then for ordered discrete and continuous choice.

4.3 Discrete choice

4.3.1 The Daly-Zachary-Williams theorem

An important early result for welfare analysis in McFadden type discrete choice models is the Daly-Zachary-Williams theorem (McFadden 1981 section 5.8, Daly-Zachary 1978, and Williams 1977). It can be viewed as the analog of Roy's identity for multinomial choice. The theorem establishes the relation between choice probabilities and mean indirect utility in discrete choice models with additively separable unobservables. In the set up laid out above, if $U_j\left(y-p_j,\eta\right)=V_j\left(y-p_j\right)+\eta_j$, $j=0,1,...,J$, so that

$$W^A\left(\mathbf{p},y,\boldsymbol{\eta}\right)=\max_{j=0,1,\ldots,J}\left\{V_j\left(y-p_j\right)+\eta_j\right\},$$

then the Daly-Zachary-Williams theorem says that for each $j=0,1,...,J$,

$$q_j\left(\mathbf{p},y\right)=\frac{\partial}{\partial\left(y-p_j\right)}E_{\boldsymbol{\eta}}\left\{W^A\left(\mathbf{p},y,\boldsymbol{\eta}\right)\right\}. \tag{24}$$

The additive structure is restrictive and has testable implications, as shown in Bhattacharya (2021). Besides, equation (24) does not tell us anything about the *distribution* of $W^A\left(\mathbf{p},y,\boldsymbol{\eta}\right)$. Therefore, even if we restrict unobserved heterogeneity to be additively separable, equation (24) cannot be used to calculate functionals other than the mean indirect utility; in particular, one cannot learn the quantiles of indirect utility at a given price.

4.3.2 Bhattacharya-Komarova (2022)

In a recent paper, Bhattacharya-Komarova (2022) have shown how to conduct policy evaluation by reconciling the utility based cost-benefit analysis tradition of public finance with the nonparametric program evaluation tradition in econometric that allows for unrestricted preference heterogeneity. They show that the distribution (and hence average, quantiles etc.) of consumers' indirect utilities can be nonparametrically identified from cross-sectional data on individual choice without functional form assumptions on utilities, preference heterogeneity and income-effects.[6]. This enables fully nonparametric evaluation of interventions and design of optimal treatment choice based on aggregate utility, in line with traditional cost-benefit analysis in public and welfare economics. Knowledge of the indirect utility distribution also permits measurement of the efficiency loss required to ensure a desired average outcome.

To see the key results formally, consider the common situation of multinomial choice outlined in Section 2.2.1 above with $J+1$ alternatives, with the outside alternative denoted by option 0. Recall that the indirect utility function is given by

$$W(\mathbf{p}, y, \eta) = \max\left\{U_0(y, \eta), U_1(y - p_1, \eta), ..., U_J(y - p_J, \eta)\right\}.$$

Since a monotone transformation of a utility function represents the same ordinal preferences and therefore leads to the same choice, we need to normalize one of the constituent alternative specific utility functions in order to give empirical content to the indirect utility function. Toward that end, suppose that for each η, the function $U_0(\cdot, \eta)$ is strictly increasing (non satiated) and continuous in the numeraire and, therefore, invertible. Then $V_0(y, \eta) \equiv y$ and $V_j(y - p_j, \eta) \equiv U_0^{-1}(U_j(y - p_j, \eta), \eta)$ is an equivalent normalization of utilities representing exactly the same individual preferences as $\{U_j(y - p_j, \eta)\}$, $j = 1, ..., J$ and $U_0(y, \eta)$. This normalization is analogous to the convention in empirical IO where utility from the outside good – alternative 0 here – is normalized to zero. The above normalization also converts the indirect utility to a money metric, thereby facilitating comparison with the cost of policy interventions. In this set-up, the marginal distribution of indirect utility induced by the distribution of η at fixed values of $\mathbf{p}$ and y is nonparametrically identified from the (structural) average demand function. To see this, note that

$$W(\mathbf{p}, y, \eta) = \max\left\{y, U_0^{-1}(U_1(y - p_1, \eta), \eta), ..., U_0^{-1}(U_J(y - p_J, \eta), \eta)\right\}, \tag{25}$$

so we have that $W(\mathbf{p}, y, \eta) \geq y$ a.s. For $c > y$, we get that

$$\begin{aligned}
& \Pr\left[\max\left\{U_0^{-1}(U_1(y - p_1, \eta), \eta), ..., U_0^{-1}(U_J(y - p_J, \eta), \eta)\right\} \leq c\right] \\
= \; & \Pr\left[\max\left\{U_1(y - p_1, \eta), ..., U_J(y - p_J, \eta)\right\} \leq U_0(c, \eta)\right] \\
= \; & \Pr\left[\max\left\{U_1(c - (c - y + p_1), \eta), ..., U_J(c - (c - y + p_J), \eta)\right\} \leq U_0(c, \eta)\right] \\
= \; & q_0(c - y + p_1, ..c - y + p_J, c).
\end{aligned}$$

Therefore, the CDF. of $W(\mathbf{p}, y, \eta)$ generated by randomness in η is given by

$$F_{W(\mathbf{p},y,\eta)}(c) \equiv \Pr\left[W(\mathbf{p}, y, \eta) \leq c\right] = \begin{cases} 0 \text{ if } c < y, \\ q_0(c - y + p_1, ..c - y + p_J, c) \text{ if } c \geq y. \end{cases} \tag{26}$$

[6]The target individuals of the intervention would generically differ along both observable and unobservable (to the researcher) determinants of preferences, and the 'distribution' of indirect utility here refers to the distribution induced by the unobservables for fixed values of observables.

The calculation of $\bar{W}^{\varepsilon}(\mathbf{p}, y)$ from (23) is facilitated by the observation that

$$\int_b^{\infty} x^{\alpha} f_X(x)\, dx = b^{\alpha}(1 - F_X(b)) + \alpha \int_b^{\infty} x^{\alpha-1}(1 - F_X(x))\, dx \tag{27}$$

using integration by parts. It follows from (26) and (27) that $\mathcal{W}^{\varepsilon}(\mathbf{p}, y)$ equals

$$\begin{aligned} & \frac{y^{1-\varepsilon}}{1-\varepsilon} \times \Pr[W(\mathbf{p}, y, \eta) = y] + \int_y^{\infty} \frac{c^{1-\varepsilon}}{1-\varepsilon} \times f_{W(\mathbf{p},y,\eta)}(c)\, dc \\ = \; & \frac{y^{1-\varepsilon}}{1-\varepsilon} + \int_0^{\infty} (z+y)^{-\varepsilon} \times \{1 - q_0(z+p_1, z+p_2, ..., z+p_J, z+y)\}\, dz.^{7} \end{aligned} \tag{28}$$

In particular, for $\varepsilon = 0$, i.e. utilitarian planner preferences, and using $E(X) = \int_0^{\infty}(1 - F(x))\, dx - \int_{-\infty}^{0} F(x)\, dx$, the average indirect utility reduces to the line integral

$$\bar{W}^{0}(\mathbf{p}, y) = y + \int_0^{\infty} \{1 - q_0(z+p_1, z+p_2, ..., z+p_J, z+y)\}\, dz. \tag{29}$$

In empirical IO, the widely used log sum measure of consumer welfare is precisely the average social welfare in the multinomial logit model, see Train (2003), Sec 3.5.

4.3.3 Binary choice

It is useful to see explicit formulae for the most common case of discrete choice, namely binary choice where $J = 1$. Consider the case where a subsidy on alternative 1 changes its price from $\bar{p}$ to $\bar{p} - \sigma$. In this case, the subsidy induced **change in average social welfare** at income y for a generic $\varepsilon \in [0, 1)$ is given by

$$\Delta(\bar{p}, \sigma, y; \varepsilon) \equiv \int_0^{\infty} (z+y)^{-\varepsilon} \times [q_1(\bar{p} - \sigma + z, y+z) - q_1(\bar{p} + z, y+z)]\, dz; \tag{30}$$

under $\varepsilon = 0$, we have that

$$\Delta(\bar{p}, \sigma, y; 0) \equiv \int_0^{\infty} [q_1(\bar{p} - \sigma + z, y+z) - q_1(\bar{p} + z, y+z)]\, dz; \tag{31}$$

the **average CV** at income y (see Bhattacharya 2015, eqn. 10) equals

$$S(\bar{p}, \sigma, y) \equiv \int_{\bar{p}-\sigma}^{\bar{p}} q_1(p, y + p - \bar{p} + \sigma)\, dp \overset{\text{subs } z = p - \bar{p} + \sigma}{=} \int_0^{\sigma} q_1(\bar{p} - \sigma + z, y + z)\, dz. \tag{32}$$

Furthermore,

$$\int_0^{\infty} q_1(\bar{p} + z, y+z)\, dz \overset{\text{sub } t = z + \sigma}{=} \int_{\sigma}^{\infty} q_1(\bar{p} + t - \sigma, y + t - \sigma)\, dt, \tag{33}$$

and therefore, from (31), (32) and (33) we have that

$$\begin{aligned} & \Delta(\bar{p}, \sigma, y; 0) - S(\bar{p}, \sigma, y) \\ = \; & \int_{\sigma}^{\infty} [q_1(\bar{p} - \sigma + z, y+z) - q_1(\bar{p} - \sigma + z, y - \sigma + z)]\, dz. \end{aligned} \tag{34}$$

[7] For standard parametric CDFs like probit or logit, the integral is bounded for $0 \leq \varepsilon \leq 1$.

Now, the integrand in (34) is strictly positive (negative) for all z if option 1 is normal (resp. inferior). Therefore, the only way $\Delta(\bar{p}, \sigma, y; 0) = S(\bar{p}, \sigma, y)$ is that $q_1(p, y)$ does not depend on y, which implies utilities are quasilinear, and therefore by (29), the social marginal utility of income equals 1.

The **average treatment effect**, the quantity most commonly used in program evaluation and the treatment choice literature, equals

$$T(\bar{p}, \sigma, y) \equiv q_1(\bar{p} - \sigma, y) - q_1(\bar{p}, y), \tag{35}$$

which is simply the integrand of (31) evaluated at the lower limit of the integral. Since this is measured as *quantity* of demand, a direct comparison with average or marginal subsidy cost is not possible. In contrast, the quantities $\Delta(\cdot, \cdot, \cdot)$ or $S(\cdot, \cdot, \cdot)$ provide theoretically justified monetary values of the choice, based on the choice makers' own preference.

Deadweight loss: The average cost of the subsidy equals $\sigma \times q_1(\bar{p} - \sigma, y)$ in every case. Therefore, the deadweight loss of the subsidy under $\varepsilon = 0$ is given by

$$\begin{aligned} & \sigma \times q_1(\bar{p} - \sigma, y) - \Delta(\bar{p}, \sigma, y; 0) \\ = & \int_{\bar{p}-\sigma}^{\bar{p}} [q_1(\bar{p} - \sigma, y) - q_1(t, y + t - \bar{p} + \sigma)]\, dt \\ & + \int_{\bar{p}}^{\infty} [q_1(t, y + t - \bar{p}) - q_1(t, y + t - \bar{p} + \sigma)]\, dt. \end{aligned} \tag{36}$$

Note that the first term in (36) is positive because

$$\begin{aligned} & q_1(t, y + t - \bar{p} + \sigma) - q_1(\bar{p} - \sigma, y) \\ = & \Pr[U_1(y - \bar{p} + \sigma, \eta) \geq U_0(y + t - \bar{p} + \sigma, \eta)] - \Pr[U_1(y - \bar{p} + \sigma, \eta) \geq U_0(y, \eta)] \\ \leq & \ 0, \text{ for } t \geq \bar{p} - \sigma \text{ since } U_0(\cdot, \eta) \text{ is strictly increasing.} \end{aligned} \tag{37}$$

The second term will be negative if the good is normal, and the deadweight loss may be *negative* if the income effect is strongly positive. This is in contrast to the deadweight loss based on the CV which must necessarily be nonnegative.

Finally, Hendren-Finkelstein's MVPF at the status-quo ($p = \bar{p}$, $\sigma = 0$) equals

$$\left. \frac{\frac{\partial}{\partial \sigma} \Delta(\bar{p}, \sigma, y; 0)}{\frac{\partial}{\partial \sigma} \{\sigma \times q_1(\bar{p} - \sigma, y)\}} \right|_{\sigma=0} \overset{\text{by (31)}}{=} \frac{-\int_0^\infty \frac{\partial q_1(\bar{p}+z, y+z)}{\partial p} dz}{q_1(\bar{p}, y)}. \tag{38}$$

4.3.4 Optimal allocation problem

The optimal subsidy targeting problem, which maximizes aggregate welfare subject to a budget constraint on subsidy spending, can be stated as follows. Suppose in the above setup, the planner is considering subsidizing alternative 1. Let M denote the planner's aggregate subsidy budget, expressed in per capita terms, $F_Y(\cdot)$ the marginal distribution of income in the population, $\sigma(y)$ the amount of subsidy that a household with income y will be entitled to, $\mathcal{T}$ denote the set of politically/practically feasible targeting rules $\sigma(\cdot)$, and $C(y,\sigma(y)) = \int \sigma(y) \times q_1(\bar{p}-\sigma(y), y)\, dF_Y(y)$ is the cost per capita of offering subsidy $\sigma(y)$ to individuals whose income is y. Then the optimal subsidy problem solves

$$\arg\max_{\sigma(\cdot)\in\mathcal{T}} \int \mathcal{W}\left(\bar{p}-\sigma(y), \mathbf{p}_{-1}, y\right) dF_Y(y) \text{ s.t. } \int C(y,\sigma(y))\, dF_Y(y) = M. \tag{39}$$

Taxes can be incorporated into the analysis by allowing $\mathcal{T}$ to contain functions that take on negative values. In particular, a revenue-neutral welfare maximizing rule that taxes the rich, i.e. $\sigma(\cdot) < 0$ and subsidizes the poor i.e. $\sigma(\cdot) > 0$, will solve (39) with $M = 0$. Bhattacharya-Komarova (2022) discuss applications of these ideas to an observational dataset involving taxes and subsidies on private tuition in India. They show that the path of the optimal subsidy as a function of income differs widely, depending on whether it is the average take-up (demand) or the aggregate social welfare that is being maximized.

Parameter uncertainty: The above formulation takes the observed individuals in the dataset as the target of the policy. If however, we consider the unobserved population from which the sample was drawn as the true target, then parameter uncertainty will need to be accounted for in our decision problem. This can be achieved by defining the loss function

$$\begin{aligned} L(\sigma(\cdot),\theta,c) &= -\int B(\bar{p},\sigma(y),y,\theta_1)\, dF(y,\theta_2) \\ &\quad + c\left[M - \int \sigma(y)\times \bar{q}_1(\bar{p}-\sigma(y),y,\theta_1)\, dF(y,\theta_2)\right]^2, \end{aligned}$$

with penalty $c > 0$ and where $\theta = (\theta_1, \theta_2)$ denotes the parameters determining the demand function (θ_1) and the marginal distribution of income (θ_2). Then define the optimal choice of $\sigma(\cdot)$ by solving

$$\min_{\sigma(\cdot)} \int L(\sigma(\cdot),\theta,c)\, dP_{post}(\theta|data), \tag{40}$$

where $P_{post}(\theta|data)$ refers to the posterior distribution of θ given the data. For computational simplicity, one can use the bootstrap distribution of θ to approximate the posterior corresponding to a flat prior (see Hastie et al. 2009).

4.3.5 Connection with Hicksian compensation

A final point here concerns the connection between Hicksian compensation based welfare analysis discussed in Section 2 and the average social welfare based approach discussed in the present section. Bhattacharya-Komarova (2022) have shown that the indirect utility function at price $\mathbf{p}$ and income y is identical to the compensated income $CV(y,\mathbf{p},\infty_J,\eta)$ that equates the individual utility at income y when none of the alternatives $1,...,J$ was available to the utility when they become available at price $\mathbf{p}$. Therefore, the average social welfare $\bar{W}^{\varepsilon}(\mathbf{p},y)$, defined in Eqn. (23)

above, is nothing but the Atkinson index of income inequality using the corresponding *compensated* – instead of ordinary – income. Furthermore, the difference in individual indirect utility between two prices $\mathbf{p}^0$ and $\mathbf{p}^1$ equals

$$W\left(\mathbf{p}^1, y, \eta\right) - W\left(\mathbf{p}^0, y, \eta\right) = CV\left(y, \mathbf{p}^1, \infty_J, \eta\right) - CV\left(y, \mathbf{p}^0, \infty_J, \eta\right) \neq -CV\left(y, \mathbf{p}^0, \mathbf{p}^1, \eta\right), \quad (41)$$

with equality holding only when there are no income-effects of price changes. Thus, in presence of income-effects, the Hicksian and the aggregate social welfare based approaches are distinct and can have contradictory policy recommendations both in regard to measuring the effectiveness of specific policy-interventions and designing optimal targeting of such interventions.

4.4 Ordered discrete choice and the continuous case

A result analogous to Theorem 1 does not hold for consumption of continuous goods such as gasoline (Poterba 1991) and food (Kochar 2005). To see why, consider the situation of ordered choice with 0 denoting the outside good and 1, 2 with unit price p denoting the two inside good (e.g. no apple, 1 apple and 2 apples). Let the utilities be $U_0(y, \eta)$, $U_1(y - p, \eta)$, $U_2(y - 2p, \eta)$. As above, normalize

$$W(p, y, \eta) \equiv \max\left\{y, U_0^{-1}\left(U_1(y-p,\eta), \eta\right), U_0^{-1}\left(U_2(y-2p,\eta), \eta\right)\right\}.$$

Now, for $c \geq y$, we have that

$$\begin{aligned}
& \Pr\left[\max\left\{y, U_0^{-1}\left(U_1(y-p,\eta), \eta\right), U_0^{-1}\left(U_2(y-2p,\eta), \eta\right)\right\} \leq c\right] \\
= \; & \Pr\left[\max\left\{U_1(y-p,\eta), U_2(y-2p,\eta)\right\} \leq U_0(c,\eta)\right] \\
= \; & \Pr\left[\max\left\{U_1(c-(c-y+p),\eta), U_1(c-(c-y+2p),\eta)\right\} \leq U_0(c,\eta)\right] \\
= \; & q_0(c-y+p, c-y+2p, c). \qquad (42)
\end{aligned}$$

But $q_0(c-y+p, c-y+2p, c)$ cannot be estimated, no matter how much p and y vary, because the data can only identify demand when the price of 2 units is twice the price of 1 unit; but $c - y + 2p \neq 2(c - y + p)$. One can however obtain bounds on (42) via

$$\begin{aligned}
& L(c; p, y) \\
\equiv \; & \max\left\{q_0(\tilde{p}, 2\tilde{p}, \tilde{y}) : \tilde{y} - \tilde{p} \geq y - p, \tilde{y} - 2\tilde{p} \geq y - 2p, \tilde{y} \leq c\right\} \\
\leq \; & \Pr\left[\max\left\{U_1(y-p,\eta), U_2(y-2p,\eta)\right\} \leq U_0(c,\eta)\right] \\
\equiv \; & q_0(c-y+p, c-y+2p, c) \\
\leq \; & \min\left\{q_0(\tilde{p}, 2\tilde{p}, \tilde{y}) : \tilde{y} - \tilde{p} \leq y - p, \tilde{y} - 2\tilde{p} \leq y - 2p, \tilde{y} \geq c\right\} \\
\equiv \; & H(c; p, y).
\end{aligned}$$

and $L(c; p, y)$ and $H(c; p, y)$ are both potentially identifiable because they represent demand in situations where price of option 2 is twice the price of option 1. Note that $L(\cdot; p, y)$ and $H(\cdot; p, y)$ satisfy all properties of CDF's.

The continuous case can be thought of as the limiting case of ordered choice, e.g. one has to pay twice as much for 2 gallons of gasoline as for 1 gallon, and by the same logic as above, the welfare distribution for this case cannot be point-identified. Indeed, in the continuous choice case, the key object of interest is the indirect utility function $V(p, y, \eta)$ in (22), whose marginal distribution cannot be point-identified from the marginal distribution of demand, as explicitly discussed in Hausman-Newey (2016). However, one can make some progress by imposing more structure on preferences. For example, suppose we posit that the demand function takes a flexible random coefficient form, e.g. $q(p, y, \eta) = \eta_0 + \eta_1 p + \eta_2 y$, where $\eta \equiv (\eta_0, \eta_1, \eta_2)$ are random coefficients distributed independently of price and income. Then, using standard results on random coefficient models (Beran-Hall 1992, Hoderlein et al. 2010), the distribution of η can be identified from the observed distribution of $q(p, y, \eta) = \eta_0 + \eta_1 p + \eta_2 y$ under sufficient independent variation of price and income in the data and appropriate regularity conditions. Now, one can solve the PDE (22) via the method of characteristics (Courant-Hilbert 2008, Chapter I.5 and II.2), for which the characteristic ordinary differential equation is given by

$$\frac{dy}{dp} = \eta_0 + \eta_1 p + \eta_2 y,$$

yielding the eventual solution (Hausman 1981 Eqn. 16)

$$W(p, y, \eta) = e^{-\eta_2 p}\left[y + \frac{1}{\eta_2}\left(\eta_1 p + \frac{\eta_1}{\eta_2} + \eta_0\right)\right].$$

Now, the joint distribution of $\eta \equiv (\eta_0, \eta_1, \eta_2)$ obtained above will yield, via the standard delta method, the marginal distribution of the indirect utility function $W(p, y, \eta)$ at fixed p and y. Once we obtain this marginal distribution, we can carry out exercises like those outlined above in this section.

5 Social Interactions

Social interaction models – where an individual's payoff from an action depends on the expected fraction of her peers choosing the same action – feature prominently in economic and sociological research. It is well-understood that in such settings targeted policies will have spillover effects, but formal welfare analysis of such policies has received little attention in the literature. This is somewhat surprising since common public interventions such as taxes and subsidies are often motivated by efficiency losses resulting from externalities.

Brock-Durlauf (2001), in their seminal paper on discrete choice under social interactions, discussed how to rank-order multiple equilibria resulting from policy interventions in terms of their aggregate, utilitarian social utility. Recently, Bhattacharya-Dupas-Kanaya (2021) have obtained closed-form expressions for average welfare effects of price changes in the setup of Brock-Durlauf (2001). In this setup, the choice probability at price p income y and neighborhood adoption rate π is given by $q(p, y, \pi) \equiv F(c_0 + c_1 p + c_2 y + \alpha\pi)$ where $F(\cdot)$ is a distribution function, and under suitable restrictions, the data can identify the the coefficients c_0, c_1, c_2, α and $F(\cdot)$. Now, for a hypothetical price change from p_0 to p_1, e.g. due to a public subsidy, Bhattacharya et al. (2021) show that values of c_0, c_1, c_2, α and $F(\cdot)$, which would imply a unique counterfactual demand prediction, in general imply a range of different welfare predictions. For example, in a model for adoption of an insecticide treated, antimalarial mosquito net (Dupas 2014), α reflects the aggregate effect of several distinct mechanisms, namely (a) a social preference for conforming, (b) learning from others' experiences, (c) a health concern led desire to protect oneself from mosquitoes deflected from neighbors who adopt a bednet, and (d) desire to free ride on other users who increase herd immunity via the insecticide effects. These distinct mechanisms are not separately identifiable from choice data (only their net effect is). But these distinct mechanisms have different implications for welfare if, say, a subsidy is introduced. In particular, suppose all spillover is due to conforming and there is no health externality. Then as more neighbors buy, a household's perceived utility from buying will increase. Conversely, suppose spillover is only due to negative health externality of buyers on non buyers. Then increased purchase by neighbors would lower the utility of a household upon *not* buying via the health route, but not affect it upon buying since the household is then protected anyway. These different welfare effects are all consistent with the same positive α.

A similar issue arises in evaluating merit-based school vouchers for attending a high achieving school, where net welfare could be negative or positive depending on whether for a potential mover, the gain from attending a cheap school with better peers dominates the loss from not moving and suffering from the quality decline in the non selective school. An aggregate positive social interaction coefficient in an individual school choice model cannot tell these two mechanisms apart.

More formally, consider a household with income y deciding whether to buy one unit of an indivisible good at price p, and believes that a fraction π of its neighbors buy the good. Utilities from not buying and buying are given by

$$\begin{aligned} U_0(y, \pi, \boldsymbol{\eta}) &= \delta_0 + \beta_0 y + \alpha_0 \pi + \eta^0, \\ U_1(y-p, \pi, \boldsymbol{\eta}) &= \delta_1 + \beta_1 (y-p) + \alpha_1 \pi + \eta^1, \end{aligned} \tag{43}$$

where we assume that $\beta_0 > 0$, $\beta_1 > 0$, i.e., non satiation in numeraire, and β_1 need not equal β_0, i.e. income-effects can be present. Then the structural choice probability of buying at (p, y, π) is given by

$$q_1(p, y, \pi) = F(\underbrace{c_0}_{\delta_1 - \delta_0} + \underbrace{c_1}_{-\beta_1} p + \underbrace{c_2}_{\beta_1 - \beta_0} y + \underbrace{\alpha}_{\alpha_1 - \alpha_0} \pi), \tag{44}$$

where $F(\cdot)$ denotes the marginal distribution function of $\eta^0 - \eta^1$. It is known from Brock-Durlauf (2007) that the structural choice probabilities $F(c_0 + c_1 p + c_2 y + \alpha\pi)$ identify c_0, c_1, c_2 and α, i.e. $(\delta_1 - \delta_0)$, β_0, β_1 and $(\alpha_1 - \alpha_0)$ up to scale even without knowledge of the probability distribution of $\eta^0 - \eta^1$.

Starting with a situation where the price of the product is p_0 and the value of π is π_0, suppose a price subsidy is introduced such that individuals with income less than a threshold τ become eligible to buy the product at price $p_1 < p_0$. This policy will alter the equilibrium adoption rate; suppose the new equilibrium adoption rate changes to π_1, with π_0 and π_1 solving the fixed point conditions:

$$\pi_0 = \int F\left(c_0 + c_1 p_0 + c_2 y + \alpha \pi_0\right) dF_Y\left(y\right), \tag{45}$$

$$\pi_1 = \int \left[1\left\{y \le \tau\right\} F\left(c_0 + c_1 p_1 + c_2 y + \alpha \pi_1\right) + 1\left\{y > \tau\right\} F\left(c_0 + c_1 p_0 + c_2 y + \alpha \pi_1\right)\right] dF_Y\left(y\right), \tag{46}$$

where F_Y is the CDF of Y.

For given values of π_0 and π_1, the compensating variation (CV) for a subsidy *eligible* individual, for any potential value of π_1 corresponding to the new equilibrium, is the solution S to the equation

$$\max\left\{U_1\left(y + S - p_1, \pi_1, \boldsymbol{\eta}\right), U_0\left(y + S, \pi_1, \boldsymbol{\eta}\right)\right\} = \max\left\{U_1\left(y - p_0, \pi_0, \boldsymbol{\eta}\right), U_0\left(y, \pi_0, \boldsymbol{\eta}\right)\right\}. \tag{47}$$

Then Bhattacharya et al. (2021) show that if $\alpha_1 \ge 0 \ge \alpha_0$, then the average compensating variation is given by

$$\begin{aligned} &\underbrace{-\int_{p_1 - p_0 - \frac{\alpha_1}{\beta_1}\left(\pi_1 - \pi_0\right)}^{0} q_1\left(p_1 - a, y, \pi_0 + \frac{\alpha_1}{\alpha}\left(\pi_1 - \pi_0\right)\right) da}_{\text{welfare gain from subsidy plus conforming effect}} \\ &+ \underbrace{\int_0^{\frac{\alpha - \alpha_1}{\beta_0}\left(\pi_1 - \pi_0\right)} \left[1 - q_1\left(p_1 - a, y, \pi_0 + \frac{\alpha_1}{\alpha}\left(\pi_1 - \pi_0\right)\right)\right] da}_{\text{welfare loss from negative externalities}}; \end{aligned} \tag{48}$$

and if $\alpha_1 \ge \alpha_0 \ge 0$, it is given by

$$\begin{cases} C_1\left(\alpha_1\right), & \text{if } \alpha \le \alpha_1 \le \frac{\beta_1}{\beta}\left(\beta_1 - \beta\right)\frac{p_0 - p_1}{\pi_1 - \pi_0} + \frac{\beta_1}{\beta}\alpha, \\ C_2\left(\alpha_1\right), & \text{if } \alpha_1 > \frac{\beta_1}{\beta}\left(\beta_1 - \beta\right)\frac{p_0 - p_1}{\pi_1 - \pi_0} + \frac{\beta_1}{\beta}\alpha. \end{cases} \tag{49}$$

where

$$C_1\left(\alpha_1\right) := -\int_{p_1 - \frac{\alpha - \alpha_1}{\beta_0}\left(\pi_1 - \pi_0\right)}^{p_0 + \frac{\alpha_1}{\beta_1}\left(\pi_1 - \pi_0\right)} q_1\left(p, y, \pi_0 + \frac{\alpha_1}{\alpha}\left(\pi_1 - \pi_0\right)\right) dp,$$

$$C_2\left(\alpha_1\right) := -\int_{p_0 + \frac{\alpha - \alpha_1}{\beta_0}\left(\pi_1 - \pi_0\right)}^{p_1 - \frac{\alpha_1}{\beta_1}\left(\pi_1 - \pi_0\right)} \left[1 - q_1\left(p, y + p - p_0, \pi_1 - \frac{\alpha_1}{\alpha}\left(\pi_1 - \pi_0\right)\right)\right] dp.$$

In both these expressions, the terms α_0, α_1 appear separately but what the choice probabilities identify is $\alpha = \alpha_1 - \alpha_0$. This leads to nonidentification of the average welfare (48) and (49). However, knowledge of α provides bounds on the range of α_1 which, together with the point-identified β_0, β_1 lead to bounds on the average welfare, e.g. for $\alpha_1 \ge 0 \ge \alpha_0$, we must have $\alpha_1 = \alpha - \alpha_0 \in [\alpha, \infty)$ etc. In particular, given α, the welfare gain in expression (48) is increasing in α_1; i.e., the welfare gain is largest in absolute value when $\alpha_1 = \alpha$ and $\alpha_0 = 0$, and the smallest when $\alpha_1 = 0$ and $\alpha_0 = -\alpha$; conversely for welfare loss. Intuitively, if there is no negative externality from increased π on non-purchasers, then they do not suffer any welfare loss, but purchasers have a welfare gain from both lower price and higher π. Conversely, if all the spillover is negative, then purchasers still receive a welfare gain via price reduction, but non-purchasers suffer welfare loss due to increased π. Analogous arguments apply in the $\alpha_1 \geq \alpha_0 \geq 0$ case. A similar set of expressions holds for welfare of individuals who are not eligible for the subsidy. The width of the bounds obtained by varying α_1 depends on the extent to which $q_1(\cdot,\cdot,\pi)$ is affected by π, i.e. the extent of social spillover, and also the difference in the realized values π_1 and π_0. Note, however, that for our single-index model, the fixed point restrictions imply that the counterfactual π_1 and π_0 depend on α_1 and α_0 only via $\alpha = \alpha_1 - \alpha_0$ which is point-identified; thus every potential value of counterfactual *demand* is point-identified, although the welfare distributions are not.

Social welfare: The underidentification result continues to hold if we focus on social, as opposed to individual welfare, as defined in Section 4. This is because the indirect utility function equals

$$W(p,y,\pi,\eta) = \max\left\{\delta_0 + \beta_0 y + \alpha_0 \pi + \eta^0, \delta_1 + \beta_1 (y-p) + \alpha_1 \pi + \eta^1\right\},$$

so that the cumulative distribution function of $W(p,y,\pi,\eta)$ is given by

$$\begin{aligned} & \Pr\left[\max\left\{\delta_0 + \beta_0 y + \alpha_0 \pi + \eta^0, \delta_1 + \beta_1 (y-p) + \alpha_1 \pi + \eta^1\right\} \leq c\right] \\ = & \Pr\left[\delta_0 + \beta_0 y + \alpha_0 \pi + \eta^0 \leq c, \delta_1 + \beta_1 (y-p) + \alpha_1 \pi + \eta^1 \leq c\right] \\ = & \Pr\left[\begin{array}{l} \eta^0 \leq c - \delta_0 - \beta_0 y - \alpha_0 \pi, \\ \eta^1 \leq c - \delta_1 - \beta_1 (y-p) - \alpha_1 \pi \end{array}\right]. \end{aligned}$$

This cannot be identified because α_0 and α_1 cannot be learnt from the identified quantity $\alpha = \alpha_1 - \alpha_0$, as above. Therefore, even if one can point identify the δs and βs and is willing to specify the joint distribution of ηs, the CDF of $W(p,y,\pi,\eta)$ remains underidentified. Furthermore, unlike the Hicksian welfare distributions above, bounding the CDF of $W(p,y,\pi,\eta)$ over possible realizations of α_0 and α_1 consistent with the identified $\alpha = \alpha_1 - \alpha_0$ would also require specification of the joint distribution of ηs, since the choice probabilities identify only the distribution of $\eta^1 - \eta^0$ which is consistent with infinitely many joint distributions.

6 Identification and Data Issues

This section deals with issues of identification of the choice probability function itself which is the core of our welfare results above.

First, note that the results of sections 2, 3 and 4 on unilateral choice (i.e. without externalities, which is discussed in Section 5) derive fully nonparametric expressions for welfare distributions in terms of the choice probability functions q. As such, they remain valid irrespective of how q itself is identified. It follows therefore that knowledge of the underlying preference distributions is not necessary for welfare analysis in the settings discussed above; knowledge of q alone suffices. In fact, the following counterexample from Bhattacharya (2015) shows that identifiability of the preference heterogeneity distribution, or even correct specification of its dimension are not a requirement for identifiability of welfare distributions.

Example: Consider a binary setting where an individual with income Y decides whether to buy a product at price P or not. Let $\eta \equiv (\eta_1, \eta_0)$ denote the vector of unobserved heterogeneity, where η_1 affects the utility when buying and η_0 when not. Assume η 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

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

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 distribution of equivalent variation and compensating variation 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, it follows that for any p, y in the support of the data, we must have that $p - y < \eta_1 < p$, or

$$0 < \frac{p - \eta_1}{y} < 1, \text{ with probability 1.} \tag{50}$$

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

$$\begin{aligned}
q_1(p, y) &= \Pr\{y - p + \eta_1 > (1-\eta_0) y\} \\
&= \Pr\left\{\eta_0 > \frac{p - \eta_1}{y}\right\} \text{ since } y > 0 \\
&= \int_T \left(1 - \frac{p - \eta_1}{y}\right) dG(\eta_1), \text{ by } \eta_1 \perp \eta_0, \text{ inequality (50) and } \eta_0 \sim U(0,1) \\
&= \left(1 - \frac{p}{y}\right) + \frac{1}{y} E(\eta_1).
\end{aligned}$$

Thus, the choice probability $q_1(p, y)$ 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, $\eta_1 \sim Uniform[p_H - y_L, 0]$ and $\eta_1 = \frac{p_H - y_L}{2}$ with probability 1 will both produce identical $q_1(p, y)$ *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 distribution of equivalent variation and compensating variation are point-identified from $q_1(\cdot, \cdot)$ as implied by theorem 1 above.

In presence of externalities, however, this attractive feature is lost, and it is no longer possible (see Section 5) to learn welfare effects without knowledge of preferences underlying the observable demand.

In what follows, we discuss four further practical issues that arise in the identification/estimation of q in real data settings.

(i) **Bounded support and shape restrictions**: Some of the welfare results described above require estimating and integrating a choice probability functional over an infinite domain. For example, the welfare effect of removing the $(J+1)$th alternative given by (12):

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

Calculating this CDF requires knowledge of $q_{J+1}(p_{10}, ..., p_{J0}, p_{J+1} + a, y + a)$ for all $a \geq 0$, which is typically not feasible in any finite dataset. This problem does not arise in parametric approaches because a few parameters guide the entire shape of the demand function, and those can be estimated from data with limited support. Regardless, it is not clear if one can trust estimates that require extrapolation well-beyond the range of the data, as would be done in the parametric approach. Continuing with the nonparametric theme, note however, that

$$q_{J+1}(p_{10}, ..., p_{J0}, p_{J+1} + a, y + a) = \int 1\left\{U_{J+1}(y - p_{J+1}, r) > \max_{k \neq J+1}\{U_k(y + a - p_{k0}, r)\}\right\} dF_\eta(r)$$

is weakly decreasing in a. Therefore, using the largest value of a (say a^U) s.t. $(p_{10}, ..., p_{J0}, p_{J+1} + a, y + a)$ being in the observed dataset gives a lower bound on the CDF for values of $a > a^U$; the upper bound being just 1. If demand takes values close to 1 at a^U, then these bounds will be tight. A similar argument applies to (11) where

$$q_{J+1}(p_{10}, ..., p_{J0}, p_{J+1} + a, y) = \int 1\left\{U_{J+1}(y - p_{J+1} - a, r) > \max_{k \neq J+1}\{U_k(y - p_{k0}, r)\}\right\} dF_\eta(r)$$

is also decreasing in a. These shape restrictions on the choice probability function can be imposed during estimation for efficiency gains (Chetverikov et al. 2018). In particular, Bhattacharya-Komarova (2022) show how to use shape preserving B-spline estimators for such shape-restricted nonparametric estimation with real data. Alternatively, one may use a flexible parametric specification via a random coefficient structure, widespread in IO applications. It takes the form

$$q_j(p_1, ..., p_J, y) = \int F_j(\gamma_1 p_1 + ... + \gamma_J p_J + \alpha y)\, dG(\boldsymbol{\gamma}, \alpha; \theta),$$

where $\boldsymbol{\gamma} = (\gamma_1 ... \gamma_J)$ and α are random variables with joint distribution $G(\cdot, \cdot; \theta)$, indexed by an unknown parameter vector θ, and $F_j(\cdot)$ is a specified CDF e.g. a probit or logit (McFadden-Train 2000, Fox et al. 2012). The typical approach in applications is to specify $G(\cdot, \cdot; \theta)$ to be multivariate normal. Having obtained estimates of θ, welfare distributions like (6), (9), (19) and (26) are all estimable via numerical integration. For example, in STATA one can estimate mixed logit models via the command "**cmmixlogit**" and perform numerical integration with the command "**integ**".

On the other hand, one can impose the shape restrictions during estimation by noting that

$$\begin{aligned}
& q_j(p_1, ..., p_J, y) \\
= & \int F_j(\gamma_1 p_1 + ... + \gamma_J p_J + \alpha y)\, dG(\boldsymbol{\gamma}, \alpha; \theta) \\
= & \int F_j\left(\left(\alpha + \sum_{j=1}^{J} \gamma_j\right) y - \gamma_1 (y - p_1) - ... - \gamma_J (y - p_J)\right) dG(\boldsymbol{\gamma}, \alpha; \theta);
\end{aligned}$$

so specifying the support of $\alpha + \sum_{j=1}^{J} \gamma_j$ and of each γ_j to lie in $(-\infty, 0)$ will be sufficient.[8]

(ii) **Interval-Income:** In many, if not most, surveys, individual income is reported in intervals. Lee-Bhattacharya (2018) show how to construct bounds on welfare in such settings, assuming the good in question is normal on average and using restrictions from economic theory. To see where such bounds come from, consider the binary choice version of (1) where good 1 is normal on average, i.e. $q_1(p, \cdot)$ is increasing, and we observe which interval $\mathcal{Y}_k \equiv [y^k, y^{k+1})$, an individual's true income lies where $k = 0, 1, ..., K-1$. Bhattacharya (2021) has recently shown that a necessary and sufficient condition for choice probabilities to have been generated from (1) are given by

$$\frac{\partial}{\partial p} q_1(p, y) + \frac{\partial}{\partial y} q_1(p, y) \leq 0 \text{ and } \frac{\partial}{\partial p} q_1(p, y) \leq 0, \text{ for any } (p, y), \tag{51}$$

which can be interpreted as the Slutsky negativity condition for binary choice. These conditions will help obtain bounds on demand and welfare predictions with interval income data. Toward that end, let $f(\cdot, \cdot)$ denote the unobserved joint density function of (P, Y), and $\Pr(p, \mathcal{Y}_k)$ denote the observed joint distribution of $(P, \mathcal{Y}_k)$. For a price increase from p_0 to p_1, the upper bound of the expected compensating variation at income y_0 equalling $\int_0^{p_1 - p_0} q_1(p_0 + a, y_0 + a)\, da$ can be obtained by solving the constrained, infinite-dimensional optimization problem:

$$\max_{f(\cdot,\cdot), q_1(\cdot,\cdot)} \int_0^{p_1 - p_0} q_1(p_0 + a, y_0 + a)\, da,$$

s.t.

$$\begin{gathered}
\int_{\mathcal{Y}_k} q_1(p, y) \times f(p, y)\, dy = \underbrace{\pi(p, \mathcal{Y}_k)\Pr(p, \mathcal{Y}_k)}_{\text{Observed}},\ k = 0, 1, ..., K, \\
\int_{\mathcal{Y}_k} f(p, y) dy = \underbrace{\Pr(p, \mathcal{Y}_k)}_{\text{Observed}},\ k = 0, 1, ..., K, \\
\frac{\partial}{\partial y} q_1(p, y) > 0, \text{ and } \frac{\partial}{\partial y} q_1(p, y) + \frac{\partial}{\partial p} q_1(p, y) < 0.
\end{gathered} \tag{52}$$

Since an admissible choice of f is where the entire probability mass within each observed income interval is concentrated at the right endpoints, we can satisfy the first constraint in (52) by

[8] For example, one can choose γ to be jointly lognormal and the conditional distribution of α given γ to be $1'\gamma - \exp\left(N\left(0, \sigma^2\right)\right)$.

setting $q_1(p,y) = q_1(p, y^{k+1})$ for all $y \in \mathcal{Y}_k$. The normal good assumption implies that the integrand $q_1(p, y_0 + p - p_0)$ in the objective function can take a value at least as high as $q_1(p, y^{k+1})$ where y^{k+1} is the right end point of the observed income interval containing $y_0 + p - p_0$. However, $q_1(p, y^{k+1})$ is also unobserved, and we will need to find an upper bound on it. Note that the probability mass on the next interval to the right, i.e. $[y^{k+1}, y^{k+2}]$ can be concentrated at y^{k+1}. Therefore, the sharp upper bound for $q_1(p, y_0 + p - p_0)$ is given by $q_1(p, [y^{k+1}, y^{k+2}))$, i.e. $\pi(p, \mathcal{Y}_{k+1})$. A similar idea works for the lower bound. Lee-Bhattacharya (2019) discuss how to operationalize the above optimization via sieve approximations to $q_1(\cdot,\cdot)$, which makes the problem equivalent to set-identifying finite dimensional parameters, namely the coefficients on the basis terms in the sieve approximation, defined via moment inequality conditions, on which there is a now an established literature in econometrics, (Andrews-Shi 2013).

(iii) **Unobserved price**: A generic problem in estimation of discrete choice models with individual data is that prices may not be observed for alternatives not chosen by the individual. One way to impute such missing price is to use the average price paid by those individuals opting for those alternatives who shop in the same location perhaps on the same day. For example, Bhattacharya-Komarova (2021) model Indian households' decision to buy private tuition for their children to supplement school education. For households that do not choose private tuition, they impute the price by using the average price of those opting for private tuition in the village or urban block of the reference household. An intuitive justification of this method is that households are likely to base their decision on the information they gather from neighbors. This method is most applicable when prices are standardized across consumers (e.g. supermarket purchase, public transport etc.) and are not individualized like wages.

(iv) **Endogenous price and income:** The third problem pertains to identification/estimation of structural choice probabilities when price or income are endogenous, conditional on other covariates. There is a significant literature in empirical IO on this problem and its potential solutions, including excellent recent surveys (Berry-Haile 2021, Gandhi-Nevo 2021). Here, we briefly summarize the main issues and point out one aspect of endogeneity that is specifically relevant to welfare analysis beyond its effect on consistent estimation of structural choice probabilities.

In discrete choice problems, price is often endogenous because it tends to be positively correlated with unobserved quality which is part of the error term. Berry-Haile (2021) discuss restrictions on demand functions and use of possible instrumental variables that enable consistent estimates of the choice probabilities in discrete choice models with market level or with individual level data. Examples include cost-shifters in the same market, price variation of the same product in other markets and non-price exogenous characteristics of competing products. Other possible techniques when individual level data are available include control function methods (Blundell-Powell 2004, Imbens-Newey 2007, Petrin-Train 2010) and special regressor methods (Lewbel 2000). In experimental settings, such as randomized allocation of vouchers and coupons, price endogeneity disappears, and the structural choice probability function can be recovered directly from the data

(Dube et al. 2018, Cohen-Dupas 2010). An analogous discussion applies to income endogeneity where unobserved preferences and observed income can be correlated conditional on observable characteristics. Note however, that when measuring welfare change due to a price-change, one can perform the analysis conditional on income. For example, consider the binary choice case and a price increase from p_0 to p_1 of alternative 1. The resulting expected equivalent variation, *conditional on* income taking an in-sample value y_0 equals $\int_0^{p_1-p_0} q_1^c\left(p_0+a|y_0\right)da$, where

$$q_1^c\left(p|y_0\right)=\int 1\left\{U_1\left(y_0-p,r\right)>U_0\left(y_0,r\right)\right\}dF_{\eta|Y}\left(r|y_0\right)$$

equals the *conditional* probability of choosing option 1 at price p among those whose current income is y_0, which differs from the unconditional probability defined in (1). But endogeneity of income itself does not affect the computation of the conditional object $\int_0^{p_1-p_0} q_1^c\left(p_0+a,y|y=y_0\right)da$, since y_0 is an in-sample value. For example, if our price variation arises from an experimental intervention where discount coupons of different values are randomly allocated among households for buying good 1, then $q_1^c\left(p,y_0|y_0\right)$ is estimable without correction for income endogeneity as long as (p,y_0) lies in the support of the data. However, this is no longer true for the average compensating variation $\int_0^{p_1-p_0} q_1^c\left(p_0+a,\underbrace{y+a}_{\neq y}|y=y_0\right)da$ because the argument of the integrand is counterfactual when income is y_0.

7 Future directions

In this final section of this survey we briefly outline five research areas where nonparametric methods of welfare analysis are yet to be fully developed in presence of unobserved heterogeneity and income-effects. These are as follows.

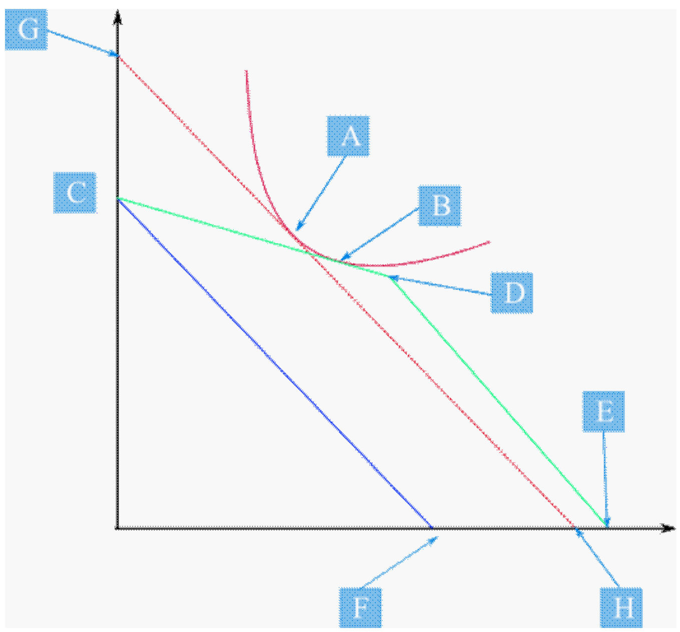
(a) Uncertainty and insurance: Provision of insurance against health and labour market risks has important welfare consequences for individuals, but can also lead to substantial deadweight loss due to under/over-provision, (Feldstein 1973, 2005). Empirical estimates of such welfare effects in the case of medical insurance were calculated in Feldman-Dowd (1991) and Manning-Marquis (1996). Manning-Marquis (1996) modelled medical expenditure following insurance, using data from the Rand health insurance experiment in the US in the 1970s where households were randomly assigned to a health plan and subsequent health events and expenses were recorded for them. The random assignment removed concerns of endogenous self-selection into health-plans, leading to consistent estimates of parametric, homogeneous indirect utility parameters from the decision to exceed the stop-loss amount, the anticipated medical expenditure (collected at baseline) and the actual expenses incurred. Based on these structural parameter estimates, they computed the trade-off between the gains from risk sharing and losses from moral hazard of the insured across a range of hypothetical insurance-plans. Einav et al. (2010a 2010b), on the other hand, calculate welfare effects in presence of adverse selection into health insurance plans offered by a firm to its employees. The key idea is that under adverse selection, insurance providers cannot distinguish between potential insurees with different health risks; consequently, realized market equilibrium is where demand equals average expected cost of insurance which happens at a higher price, lower demand situation than the socially optimal situation where demand equals the marginal cost curve. The area between these intersections is akin to the classic Harberger (Harberger 1964) triangle representing deadweight loss. Unobserved risk heterogeneity is central to Einav et al.'s analysis who allow for this heterogeneity to be entirely unrestricted, while maintaining lack of income-effects, and show that consistent estimation of demand and cost curves are sufficient for estimating the welfare loss due to information asymmetry. The Einav et al. framework would be an excellent starting point for investigating more realistic scenarios where (i) individuals are heterogeneous in terms of both risk and risk aversion, and (ii) price changes have income-effects. This framework can then be used to design optimal regulation of the insurance industry in terms of taxes/subsidies or caps on price and out-of-pocket expenses (Blomqvist 1997).

(b) Intra-household considerations: In real life, consumption decisions in multi-member households are usually made collectively. Chiappori (1992), Chiappori-Meghir (2014), and Chercheye et al. (2018) address the problem of how to evaluate individual welfare effects in such situations, using an empirically estimable structural model of household behavior, where (collective) decisions are made regarding consumption of both private and public goods. In such settings, one would typically be interested in calculating two types of welfare effects (A) the impact of policy interventions such as taxing tobacco and alcohol which are primarily consumed by male members, paying child support to women, a rise in the female member's wage relative to the male's (Chercheye et al. 2018) etc., and (B) the minimum income needed by a member living with other household members to achieve the same utility level when the she lives alone and hence pays fully for the public goods, as opposed to sharing it; such analysis informs cost of living calculations and hence the design of government transfers, inflation rates etc. How unobserved preference heterogeneity can be incorporated in such models and how they affect welfare analysis in practical settings remain to be fully investigated.

(c) Nonlinear budgets: Nonlinear pricing is common in many markets (Hausman 1985, Moffitt 1986, Wilson 1993, Armstrong 2016, and Reiss-White 2006). Supermarkets typically offer lower unit price for larger packages, regulated utilities such as electricity and water companies use peak-load pricing, and mobile phone-plans typically involve piecewise linear tariffs. The welfare impact of such pricing is of significant policy interest for regulators. A recurrent theme of research in labour and public finance is the elasticity of taxable income when marginal tax-rate is progressive, leading to nonlinear budget sets for individual workers (Saez et al. 2012, Blomquist et al. 2021). Income support schemes and the earned income tax credit also give rise to such nonlinearities.

A practically important setting of kinked budgets arise in rationed food subsidies in developing countries such as subsidized foodgrains in the Indian Public Distribution System and milk in the

Mexican NADYSRA. In these schemes, eligible households can buy up to a fixed quantity of subsidized staples at a price lower than the market price. Any additional purchase needs to be made in the open market. Such schemes produce piecewise linear but convex budget sets. However, due to various sources of leakage, corruption and possible deadweight loss, there are ongoing debates about whether such subsidies should be replaced by a universal basic income (Currie-Gahvari 2008, Khera 2014, Gentilini et al. 2020). Formally addressing this debate would require one to quantify the deadweight loss from the subsidies, to calculate the utility-preserving cash equivalent – which would inform the determination of an appropriate value of the UBI – and to measure the impact of such potential interventions on redistribution and aggregate welfare. The following graph sketches the basic idea of what such calculations entail. In this graph, the horizontal axis represents the quantity of the subsidized good, say rice, and the vertical axis represents the quantity of numeraire consumed by a household. Under the rationed subsidy, it faces the kinked budget line CDE and maximizes utility at the point A.



Once the subsidy is withdrawn and everyone faces the market price for rice, the budget line moves to CF. An income compensation at the market price would be required to move the budget line so that it becomes tangent to the original indifference curve through A, thereby allowing the consumer the same utility as with the subsidy in place. The extent of income required is the compensating variation GC. The question is how does one calculate the distribution of this compensation as the indifference curve varies across individuals and one observes the distribution of demand on the straight line segments of the budget curves and at the kink point D.

(d) **Behavioral welfare analysis**: Behavioral deviations of individuals from predictions of canonical rational choice models have been studied extensively in economics both theoretically and in a variety of practical contexts. A relatively small literature has recently developed on the problem of welfare analysis of policy interventions in such settings. Examples include (i) taxes on unhealthy food such as alcohol and soda, that aim to mitigate 'internalities' i.e. costs from suboptimal consumption (Allcott et al. 2019, Griffith et al. 2019), (ii) information disclosure and subsidies that 'nudge' consumer choice towards socially optimal outcomes (Allcott-Taubinsky 2015), and (iii) increasing the salience of existing interventions such as sales taxes, (Chetty et al. 2009). Welfare evaluation in these settings amounts to taking the observed response to an intervention to be possibly suboptimal and 'behavioral', i.e. based on the so-called "decision utility", but the welfare functional to be based on the rational or "nominal" utilities. For example, in the Chetty et al. setup (see Chetty et al. 2009 Section V), all consumers are assumed to have identical preferences over the quantity x of a continuous good and wealth y, described by the utility function $u(x)+v(y)$. The budget constraint she faces is

$$(p+\tau) \times x(p, y, \tau) + y(p, y, \tau),$$

where y denotes consumer wealth, p is the price and τ is the sales-tax which may or may not be fully salient to the consumer. Letting the consumer's optimal consumption quantity as $x^*(p, y, \tau)$ leads to the indirect utility

$$V(p, \tau, y) = u(x^*(p, y, \tau)) + v(y - (p+\tau) \times x^*(p, y, \tau)),$$

and tax revenue $\tau x^*(p, y, \tau)$. Denoting the consumer's expenditure function by $e(p, \tau, V)$, the equivalent variation of the tax equals $y - e(p, 0, V(p, \tau, y))$. Using a quadratic approximation to the resulting excess burden $y - e(p, 0, V(p, \tau, y)) - \tau x^*(p, y, \tau)$ by a Taylor expansion around $\tau = 0$ along the lines of Harberger (1964), Chetty et al. investigate theoretically the welfare impact of a *small* sales tax when taxes are fully salient ($x^*(p, y, \tau) = x^*(p+\tau, y, 0)$) and when they are not. An interesting conclusion is that lack of salience can increase or reduce consumption distortions typically associated with sales taxes depending on whether income-effects are present or absent. Taking these theoretical formulae to the data would require parametric assumptions on utilities in addition to the maintained assumption of identical tastes across consumers. Incorporating unobserved heterogeneity in preferences would make such exercises more credible since both preferences and the experience of salience are likely to vary across individuals. It is also important to examine if the assumption of identical tastes is important for identifying policy effects, e.g. taxes/subsidies, increasing the salience of taxes etc. in such behavioral models.

Allcott-Taubinsky (2015) consider a setting of binary choice between an energy efficient electric bulb and a traditional inefficient one, where heterogeneous consumers can misperceive the value of the efficient bulb. In this setting, they derive expressions to measure the demand and welfare effects of a subsidy on the energy efficient product in terms of the observed market demand curve and the 'average marginal bias' which is the mean bias across those consumers who are exactly indifferent between buying or not $E(b|vb = p)$, where v is the true value of the energy efficient bulb, p is its price and b is the misperception. This last expression is not directly observed, and the authors generate data to estimate it using an experimental nudge. As in the Chetty et al. paper, Allcott-Taubinsky (2015) formally describe the welfare trade-off achieved by a subsidy in such cases between the distortion in market consumption vis-a-vis correction of the perception bias or 'internality'.

Griffith et al. (2019), on the other hand, model consumption of different types of alcohol in Britain using a quasilinear (i.e. with no income-effects) and random coefficient BLP type utility function capturing rich consumer heterogeneity, and estimate it using individual consumption data. They then obtain a measure of the 'internalities' of alcohol drinking (which are ignored by consumers when choosing optimal consumption) by calibrating the parameters of an assumed function of alcohol consumption, based on government reports of financial costs of alcohol abuse. They use these estimates to maximize a welfare function that comprises the sum of individual indirect utilities and the tax revenue less the externalities. The solution optimally trades off the internality costs with the market distortions resulting from the tax, and conforms to the intuition that heavier drinkers should face a higher average and marginal tax rate.

(e) **Utilizing novel data sources:** With retail activity increasingly moving online, the context of welfare analysis has expanded to include the so called 'gig economy'. At the same time, the availability of large, granular data collected from websites and apps have opened up new opportunities for measuring welfare effects of economic changes in such settings. For example, Cohen et al. (2016) measures the Marshallian consumer surplus generated by UberX's cab-hailing app, using large price variation resulting from Uber's surge pricing algorithm and using detailed trip data for nearly 50 million trips in the US. Their approach is nonparametric and consists of estimating demand using a component of the price variation that results from a source of randomness that is intrinsic to Uber's price setting algorithm. In another highly innovative study using GPS based driving behavior data in Bangalore, Kreindler (2020) has estimated welfare effects of congestion pricing on driving behavior and welfare. His analysis uses a fully parametric structural model of traffic congestion, based on Arnott et al. (1993). The approach is to first estimate commuter preferences over travel time and costs using experimentally generated price variation, and then using these estimates to calculate welfare effects of congestion charges. The increasing availability of such online data of large size is likely to mitigate the computational difficulty associated with nonparametric estimation of welfare effects. Nonetheless, a large proportion of such data are unstructured, creating the scope for gainfully employing machine learning methods. For example, existing consumer reviews heavily influence future consumers' choice but appear in qualitative textual forms; machine learning methods are well-equipped to handle textual data (see Gentzkow et al. 2019). Exploring the connections and synergies between traditional economic approaches like Hicksian aggregation or hedonic i.e. characteristics-based demand models and data features revealed by machine learning methods appears to be a useful agenda for future research.

References

1. Ahmad, E. and Stern, N. 1984. "The theory of reform and Indian indirect taxes." Journal of Public Economics, 25(3): 259-298.

2. Allcott, H. and Taubinsky, D. 2015. "Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market." American Economic Review 105(8): 2501-38.

3. Allcott, H., B Lockwood, and Taubinsky, D. 2019. "Regressive sin taxes, with an application to the optimal soda tax." Quarterly Journal of Economics, 134(3): 1557–1626.

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

5. Andrews, D.W. and Shi, X. 2013. "Inference based on conditional moment inequalities." Econometrica, 81(2): 609-666.

6. Armstrong, M. 2016. "Nonlinear pricing." Annual Review of Economics, 8: 583-614.

7. Arnott, R., De Palma, A. and Lindsey, R. 1993. "A structural model of peak-period congestion: A traffic bottleneck with elastic demand." The American Economic Review: 161-179.

8. Atkinson, A.B. 1970. "On the measurement of inequality." Journal of economic theory, 2(3): 244-263.

9. Banks, J., Blundell, R. and Lewbel, A. 1996. "Tax reform and welfare measurement: do we need demand system estimation?." The Economic Journal, 106(438), 1227-1241.

10. Beran, R. and Hall, P. 1992. "Estimating coefficient distributions in random coefficient regressions." The Annals of Statistics: 1970-1984.

11. Bergson, Abram. 1954. "On the concept of social welfare." The Quarterly Journal of Economics (1954): 233-252.

12. Berry, S., Levinsohn, J. and Pakes, A. 1995. "Automobile prices in market equilibrium." Econometrica: Journal of the Econometric Society: 841-890.

13. Berry S, Haile P. 2016. Identification in differentiated products markets. Annual Review of Economics, 8: 27-52.

14. Berry, S.T. and Haile, P.A. 2021. "Foundations of demand estimation", No. w29305. National Bureau of Economic Research.

15. Bhattacharya, D. 2015. "Nonparametric welfare analysis for discrete choice." Econometrica, 83(2), 617–649.

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

17. Bhattacharya, D. 2021. "The empirical content of binary choice models." Econometrica, 89(1): 457-474.

18. Bhattacharya, D., P. Dupas and S. Kanaya 2018. "Empirical welfare analysis for discrete choice under social interactions." mimeo.

19. Bhattacharya, D. and Komarova, T. 2022. "Incorporating social welfare in program-evaluation and treatment choice." arXiv preprint arXiv:2105.08689.

20. Blackorby, C. and Donaldson, D. 1988. "Money metric utility: A harmless normalization?." Journal of Economic Theory", 46(1), 120-129.

21. Blomqvist, Å., 1997. "Optimal non-linear health insurance." Journal of health economics, 16(3): 303-321.

22. Blomquist, S., Newey, W.K., Kumar, A. and Liang, C.Y. 2021. "On bunching and identification of the taxable income elasticity." Journal of Political Economy, 129(8): 2320-2343.

23. Blundell, R. and Powell, J. 2003. "Endogeneity in nonparametric and semiparametric regression models", in Advances in Economics and Econometrics, Cambridge University Press, Cambridge, U.K.

24. Brock WA, Durlauf SN. 2001. "Discrete choice with social interactions." The Review of Economic Studies, 68(2):235-60.

25. Cherchye, L., Cosaert, S., De Rock, B., Kerstens, P.J. and Vermeulen, F. 2018. "Individual welfare analysis for collective households." Journal of Public Economics 166: 98-114.

26. Chetty, R. 2009. "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods." Annu. Rev. Econ., 1(1): 451-488.

27. Chetty, R., Looney, A. and Kroft, K. 2009. "Salience and taxation: Theory and evidence." American economic review, 99(4): 1145-77.

28. Chetverikov, D., Santos, A. and Shaikh, A.M. 2018. "The econometrics of shape restrictions." Annual Review of Economics, 10: 31-63.

29. Chiappori, P.A. 1992. "Collective labor supply and welfare." J. Polit. Econ. 100, 437–467.

30. Chiappori, P.A. and Meghir, C. 2014. "Intra-household welfare." National Bureau of Economic Research, No. w20189.

31. Cohen, J. and Dupas, P. 2010. “Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment.” Quarterly journal of Economics, 125(1)

32. Cohen, P., Hahn, R., Hall, J., Levitt, S. and Metcalfe, R. 2016. “Using big data to estimate consumer surplus: The case of uber”, (No. w22627). National Bureau of Economic Research.

33. Courant, R. and Hilbert, D. 2008. “Methods of mathematical physics: partial differential equations.” John Wiley & Sons.

34. Currie, J. and Gahvari, F. 2008. “Transfers in cash and in-kind: Theory meets the data.” Journal of Economic Literature, 46(2): 333-83.

35. 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.

36. Daly, A. and Zachary, S. 1978. “Improved multiple choice models”, in D. Henscher and Q. Dalvi (eds.) Identifying and Measuring the Determinants of Mode Choice, Teakfields, London

37. Deaton, A. 1984. “Econometric issues for tax design in developing countries”, reprinted in “The Theory of Taxation for Developing Countries.” Washington, DC: World Bank, 1987.

38. De Nardi, M., French, E. and Jones, J.B. 2016. “Medicaid insurance in old age.” American Economic Review, 106(11): 3480-3520.

39. Diamond P., Hausman J. 1994. “Contingent valuation: is some number better than no number?” Journal of Economic Perspectives. Dec; 8(4):45-64.

40. Domencich, T. and McFadden, D. 1975. “Urban travel demand - a behavioral analysis.” North-Holland, Oxford.

41. Dreze, J. 1998. “Distribution matters in cost-benefit analysis: Comment on K.A. Brekke.” Journal of Public Economics, 70(3): 485-488.

42. Dubé, J.P., Fang, Z., Fong, N. and Luo, X. 2017. “Competitive price targeting with smartphone coupons.” Marketing Science, 36(6): 944-975.

43. Dupas, P. 2014. “Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment.” Econometrica, 82(1): 197-228.

44. Einav, L., Finkelstein, A. and Cullen, M.R. 2010a. “Estimating welfare in insurance markets using variation in prices.” The quarterly journal of economics, 125(3): 877-921.

45. Einav, L., Finkelstein, A. and Levin, J. 2010b. “Beyond testing: Empirical models of insurance markets.” Annual Rev. Econ.2(1): 311-336.

46. Feldman, R. and Dowd, B. 1991. “A new estimate of the welfare loss of excess health insurance.” The American Economic Review, 81(1): 297-301.

47. Feldstein, M.S. 1973. “The welfare loss of excess health insurance.” Journal of Political Economy, 81(2, Part 1): 251-280.

48. Feldstein, M. 2005. “Rethinking social insurance.” American Economic Review, 95(1): 1-24.

49. Finkelstein, A. and Hendren, N. 2020. “Welfare analysis meets causal inference.” Journal of Economic Perspectives, 34(4): 146-67.

50. Fox, J.T., il Kim, K., Ryan, S.P. and Bajari, P. 2012. “The random coefficients logit model is identified.” Journal of Econometrics, 166(2): 204-212.

51. Gandhi, A. and Nevo, A.2021. “Empirical models of demand and supply in differentiated products industries” No. w29257. National Bureau of Economic Research.

52. Gentilini, Ugo; Grosh, Margaret; Rigolini, Jamele; Yemtsov, Ruslan. 2020. “Exploring universal basic income : A guide to navigating concepts, evidence, and practices. Washington, DC: World Bank.

53. Gentzkow, M. 2007. “Valuing new goods in a model with complementarity: Online newspapers. American Economic Review, 97(3): 713-744.

54. Gentzkow, M., Kelly, B. and Taddy, M. 2019. “Text as data. Journal of Economic Literature, 57(3): 535-74.

55. Goolsbee, A. and Petrin, A., 2004. “The consumer gains from direct broadcast satellites and the competition with cable TV. Econometrica, 72(2): 351-381.

56. Griffith, R., M. O’Connell and K. Smith 2019. “Tax design in the alcohol market, Journal of Public Economics, Vol 172, April: 20-35

57. Hammond, P.J., 1990. “Interpersonal comparisons of utility: Why and how they are and should be made, European University Institute.

https://cadmus.eui.eu/bitstream/handle/1814/342/1990_EUI%20WP_ECO_003.pdf?sequence=1

58. Hanemann, W. M. 1984. “Welfare evaluations in contingent valuation experiments with discrete responses.” American Journal of Agricultural Economics, 66(3):332-41.

59. Harberger, A.C. 1964. “The measurement of waste.” The American Economic Review, 54(3): 58-76.

60. Hastie, T., Tibshirani, R., Friedman, J.H. and Friedman, J.H. 2009. “The elements of statistical learning: data mining, inference, and prediction.” (Vol. 2: 1-758). New York: springer.

61. Hausman, J. 1981. “Exact Consumer’s Surplus and Deadweight Loss.” The American Economic Review, Vol. 71, 4, 662-676.

62. Hausman, J. and McFadden, D. 1984. “Specification tests for the multinomial logit model.” Econometrica: Journal of the econometric society: 1219-1240.

63. Hausman, J. 1985. “The econometrics of nonlinear budget sets.” Econometrica: 1255-1282.

64. Hausman J. and Newey, W. 2016. “Individual heterogeneity and average welfare.” Econometrica; 84(3):1225-48.

65. Hausman, J. and Newey, W. 2017. “Nonparametric Welfare Analysis.” Annual Review of Economics, Vol. 9: 521-546.

66. Heckman, J.J., 2010. “Building bridges between structural and program evaluation approaches to evaluating policy.” Journal of Economic literature, 48(2): 356-98.

67. Heckman, J.J., Matzkin, R.L. and Nesheim, L. 2010. “Nonparametric identification and estimation of nonadditive hedonic models.” Econometrica, 78(5): 1569-1591.

68. Heckman, J.J. and Vytlacil, E.J., 2007. “Econometric evaluation of social programs, Causal models, structural models and econometric policy evaluation.” Handbook of Econometrics, 6: 4779-4874.

69. Hendel, I. 1999. “Estimating multiple-discrete choice models: An application to computerization returns.” The Review of Economic Studies, 66(2): 423-446.

70. Herriges, J and Kling, C. 1999. “Nonlinear income effects in random utility models.” The Review of Economics and Statistics, Vol. 81, No. 1, 62-72.

71. Hicks, J.R., 1943. “The four consumer’s surpluses. The Review of Economic Studies, 11(1): 31-41.

72. Hninn, S, S. Kaneko, K. Keisuke and Yoshida, Y. 2017. “A nonparametric welfare analysis via a randomized conjoint field experiment: an application to water quality improvement and the floating settlements on Inlay Lake, Myanmar.” IDEC DP2 Series 7, no. 3 (2017): 1-33.

73. Hoderlein, S., Klemelä, J. and Mammen, E. 2010. “Analyzing the random coefficient model nonparametrically.” Econometric Theory, 26(3): 804-837.

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

75. Imbens, G.W. and Newey, W.K. 2009. “Identification and estimation of triangular simultaneous equations models without additivity.” Econometrica, 77(5): 1481-1512.

76. Jha, S. and Srinivasan, P.V. 2001. "Taking the PDS to the poor: Directions for further reform." Economic and Political Weekly: 3779-3786.

77. Kaldor, N. 1939. "Welfare propositions in economics and interpersonal comparisons of utility." Economic Journal. The Economic Journal, Vol. 49, No. 195. 49 (195): 549–552.

78. Kamat V. and Norris, S. 2021. "Estimating welfare effects in a nonparametric choice model: The Case of School Vouchers." mimeo.

79. Khera, R. 2014. "Cash vs. in-kind transfers: Indian data meets theory." Food Policy, 46: 116-128.

80. Kitagawa, T. and Tetenov, A. 2018. "Who should be treated? empirical welfare maximization methods for treatment choice." Econometrica, 86(2): 591-616.

81. Kochar, A. 2005. "Can targeted food programs improve nutrition? An empirical analysis of India's public distribution system." Economic Development and Cultural Change, 54(1): 203-235.

82. Kreindler, G. 2020. "The welfare effect of road congestion pricing: Experimental evidence and equilibrium Implications." mimeo. Harvard University

83. Kroft, K., Luo, Y., Mogstad, M. and Setzler, B.2020. "Imperfect competition and rents in labor and product markets: The case of the construction industry." (No. w27325). National Bureau of Economic Research.

84. Lee, Y.Y. and Bhattacharya, D. 2019. "Applied welfare analysis for discrete choice with interval-data on income." Journal of econometrics 211(2): 361-387.

85. Lewbel A. 2000. "Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables." J. Econometrics, 97 (1): 145-177

86. Lewbel, A. 2001. "Demand systems with and without errors." The American Economic Review, Vol. 91, No. 3.

87. Manning, W.G. and Marquis, M.S. 1996. "Health insurance: the trade-off between risk pooling and moral hazard." Journal of Health Economics, 15(5): 609-639.

88. Manski, C. F., 2004. "Statistical treatment rules for heterogeneous populations." Econometrica, 72(4): 1221-1246.

89. Manski, C. F., 2014. "Identification of income–leisure preferences and evaluation of income tax policy." Quantitative Economics, 5(1): 145-174.

90. Marschak J. 1953. "Economic measurements for policy and prediction." In: Hood W, Koopmans T, editors. Studies in Econometric Method: 1–26. Wiley, New York.

91. Mas-Colell, A., Whinston, M.D. and Green, J.R. 1995. “Microeconomic theory.” (Vol. 1). New York: Oxford university press.

92. Matzkin RL. 2007. “Heterogeneous choice.” Econometric Society Monographs. Jan; 43:75.

93. Matzkin, R.L. 2016. “On independence conditions in nonseparable models: Observable and unobservable instruments.” Journal of Econometrics, 191(2): 302-311.

94. Mayshar, J. 1990. “On measures of excess burden and their application.” Journal of Public Economics, 43(3): 263-289.

95. McFadden, D. 1973. “Conditional logit analysis of qualitative choice behavior.” elsa.berkeley.edu

96. McFadden, D. 1981.“Econometric models of probabilistic choice” Structural analysis of discrete data with econometric applications 198272 (1981).

97. McFadden D & Train K. 2000. “Mixed MNL models for discrete response.” Journal of Applied Econometrics 15: 447–470.

98. Mirrlees, J.A. 1971. “An exploration in the theory of optimum income taxation.” The Review of Economic Studies, 38(2): 175-208.

99. Moffitt, R. 1986. “The econometrics of piecewise-linear budget constraints: a survey and exposition of the maximum likelihood method.” Journal of Business & Economic Statistics, 4(3): 317-328.

100. Petrin A. 2002. “Quantifying the benefits of new products: The case of the minivan.” Journal of political Economy 110(4), 705-29.

101. Petrin, A. and Train, K.2010. “A control function approach to endogeneity in consumer choice models.” Journal of Marketing Research, 47(1): 3-13.

102. Porto, G.G., 2006. “Using survey data to assess the distributional effects of trade policy.” Journal of International Economics, 70(1): 140-160.

103. Poterba, J.M. 1991. “Is the gasoline tax regressive?. Tax policy and the economy, 5: 145-164.

104. Reiss, P.C. and White, M.W. 2006. “Evaluating welfare with nonlinear prices. National Bureau of Economic Research, No. w12370.

105. Samuelson, P.A. 1947. “Foundations of Economic Analysis, Ch. VIII, "Welfare Economics".

106. Saez, E., 2001. “Using elasticities to derive optimal income tax rates. The Review of Economic Studies, 68(1): 205-229.

107. Saez, E., Slemrod, J. and Giertz, S.H. 2012. “The elasticity of taxable income with respect to marginal tax rates: A critical review. Journal of Economic Literature, 50(1): 3-50.

108. Slesnick, D.T. 1998. “Empirical approaches to the measurement of welfare. Journal of Economic Literature, 36(4): 2108-2165.

109. Small, K. and Rosen, H. 1981. “Applied welfare economics with discrete choice models. Econometrica. Vol. 49, No. 1 105-130.

110. Small, K.A. and Hsiao, C. 1985. “Multinomial logit specification tests. International Economic Review: 619-627.

111. Trajtenberg, M. 1989. “The welfare analysis of product innovations, with an application to computed tomography scanners.” Journal of Political Economy, 97(2): 444-479.

112. Train, K. 2009. “Discrete choice methods with simulation, Cambridge University Press, Cambridge, UK.

113. Vartia, Y. O. 1983. “Efficient methods of measuring welfare change and compensated income in terms of ordinary demand functions.” Econometrica: 79-98.

114. Vytlacil, E. and Yildiz, N. 2007. “Dummy endogenous variables in weakly separable models.” Econometrica, 75(3): pp.757-779.

115. Williams, H.C. 1977. “On the formation of travel demand models and economic evaluation measures of user benefit.” Environment and planning A, 9(3): 285-344.

116. Wilson, R.B. 1993. “Nonlinear pricing.” Oxford University Press.