FOUNDATIONS OF WELFARE ECONOMICS AND PRODUCT MARKET APPLICATIONS

Daniel McFadden[1]

March 21, 2016 (revised August 20, 2017)

ABSTRACT: *A common problem in applied economics is to determine the impact on consumers of changes in prices and attributes of marketed products as a consequence of policy changes. Examples are prospective regulation of product safety and reliability, or retrospective compensation for harm from defective products or misrepresentation of product features. This paper reexamines the foundations of welfare analysis for these applications. We consider discrete product choice, and develop practical formulas that apply when discrete product demands are characterized by mixed multinomial logit models and policy changes affect hedonic attributes of products in addition to price. We show that for applications that are retrospective, or are prospective but compensating transfers are hypothetical rather than fulfilled, a Market Compensating Equivalent measure that updates Marshallian consumer surplus is more appropriate than Hicksian compensating or equivalent variations. We identify the welfare questions that can be answered in the presence of partial observability on the preferences of individual consumers. We examine the welfare calculus when the experienced-utility of consumers differs from the decision-utility that determines market demands, as the result of resolution of contingencies regarding attributes of products and interactions with consumer needs, or as the result of inconsistencies in tastes and incomplete optimizing behavior. We conclude with an illustrative application that calculates the welfare impacts of unauthorized sharing of consumer information by video streaming services.*

JEL Classification: D60, D61, D11, D12, K13, L51

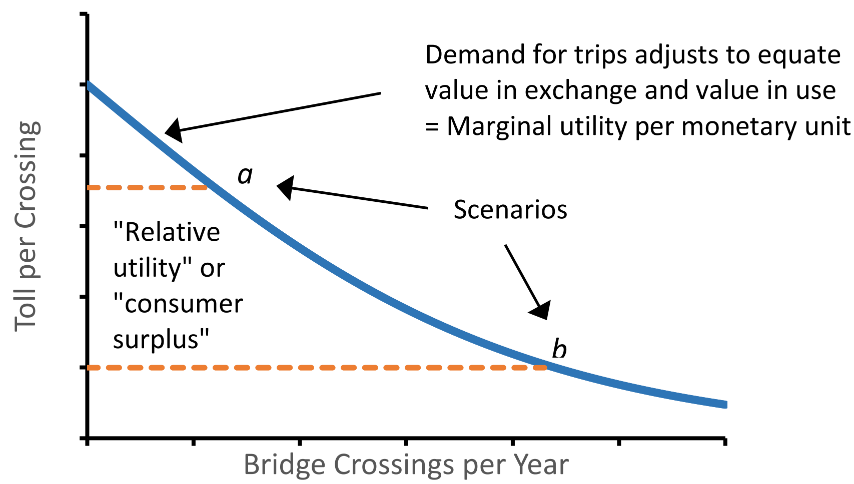
---

[1] Presidential Professor of Health Policy and Economics, University of Southern California, and E. Morris Cox Professor of Economics, University of California, Berkeley. I am indebted to Kenneth Train, Professor of Economics, University of California, Berkeley, who made major contributions to the contents of this paper, including the welfare calculus formulas given in Sections 5 and 7, the application given in Section 8, and Appendix C. I also thank Moshe Ben-Akiva, Andrew Daly, Mogens Fosgerau, Garrett Glasgow, Stephane Hess, Armando Levy, Douglas MacNair, Charles Manski, Rosa Matzkin, Kevin Murphy, Frank Pinter, Joan Walker, and Ken Wise for useful suggestions and comments.

1. INTRODUCTION

A common problem in applied economics is assessment of the welfare consequences for consumers of policies/scenarios that regulate markets for products, or correct for past product defects or misrepresentations. Examples are (1) *prospective* regulation of information provided on coverage and costs in insurance contracts and other financial instruments such as mortgages, and *retrospective* redress of harm from failures to properly disclose information; (2) harm from environmental damage to recreation facilities such as ocean beaches; (3) safety regulation of consumer products such as automobile air bags, mobile phones, and privacy protection in video streaming services, or redress of harm from safety defects; and (4) evaluation of overall market performance; e.g., the prospective benefit of blocking a merger of dominant suppliers, or retrospective harm from collusion or restraints on entry. This paper reexamines the foundations of welfare analysis for these applications, and provides a practical framework for analysis that rests on these foundations.

Figure 1. Dupuit's Calculation of Relative Utility



Measuring changes in consumer well-being from policies that affect the availability, prices, and/or attributes of goods and services has been a central concern of economics from its earliest days. Adam Smith (1776) observed that "haggling and bargaining in the market" would achieve "rough equality" between value in use and value in exchange. Working at the fringes of mainstream economics, Jules Dupuit (1844) was remarkably prescient, recognizing that if the marginal utility of income (MUI) is constant, then the demand curve for a commodity (illustrated in Figure 1) is a marginal utility curve, so that the area to the left of this demand curve between the prices established by scenarios labeled *a* and *b* gives a money-metric measure of "relative utility". Dupuit's measure later became known as *Marshallian Consumer Surplus* (MCS); see Alfred Marshall (1890, III.IV.2-8).

Hermann Gossen (1854) deduced further that consumers exhibiting diminishing marginal utility would achieve maximum utility when the marginal utilities per unit of expenditure on each good are equal, and equal the MUI.

To rephrase these propositions in current microeconomic terms, suppose the consumer maximizes a utility function $U(q_0,q_1)$ of two goods subject to a budget constraint $I = p_0q_0 + p_1q_1$, where I is income and p_0 and p_1 are the goods prices. Let $q_0 = D_0(I,p_0,p_1)$ and $q_1 = D_1(I,p_0,p_1) \equiv (I - p_0D_0(I,p_0,p_1))/p_1$ denote the demands that come out of this maximization, and let $V(I,p_0,p_1) \equiv U(D_0(I,p_0,p_1),D_1(I,p_0,p_1)) \equiv \max_{q_0} U(q_0,(I - p_0q_0)/p_1)$ denote the resulting *indirect* (or maximized) utility. The first-order condition for maximization is $FOC \equiv \partial U/\partial q_0 - (p_0/p_1)\partial U/\partial q_1 = 0$. The derivatives of V are $\partial V/\partial I = (1/p_1)\partial U/\partial q_1 + FOC\cdot(\partial D_0/\partial I) \equiv (1/p_1)\partial U/\partial q_1$ and $\partial V/\partial p_1 = -(D_1(I,p_0,p_1)/p_1)\partial U/\partial q_1 + FOC\cdot(\partial D_0/\partial p_1) \equiv -D_1(I,p_0,p_1)\cdot(\partial V/\partial I)$, illustrating the envelope theorem. Rearranging the MUI $\partial V/\partial I$ gives Smith's proposition: "value in exchange" $\equiv p_1 = (\partial U/\partial q_1)/(\partial V/\partial I) \equiv$ "value in use" (or marginal utility per unit of good 1 measured in money units), which combined with a rearrangement $(1/p_1)\,\partial U/\partial q_1 = (1/p_0)\partial U/\partial q_0$ of the FOC gives Gossen's result. The ratio ${}^{\partial V/\partial p_1}\!/_{\partial V/\partial I} \equiv -D_1(I,p_0,p_1)$ gives Roy's (1947) identity. Substituting this ratio in the Dupuit's relative utility or Marshallian consumer surplus (MCS) integral,

$$(1)\qquad MCS = \int_{p_{1b}}^{p_{1a}} D_1(I,p_0,p_1)dp_1 \equiv \int_{p_{1a}}^{p_{1b}} \frac{\partial V/\partial p_1}{\partial V/\partial I}\, dp_1 \equiv [V(I,p_0,p_{1b}) - V(I,p_0,p_{1a})]/MUI^*,$$

where the last equality is obtained by integration after applying the first mean value theorem to move outside the integral an intermediate value MUI* of the denominator $\partial V/\partial I$. In this paper, we define a measure of the consumer's change in well-being that we term the *Market Compensating Equivalent* (MCE),

$$(2)\qquad MCE = [V(I,p_0,p_{1b}) - V(I,p_0,p_{1a})]/MUI_a,$$

the difference in indirect utilities, scaled to money-metric units by dividing by the MUI at the "default" or "as is" scenario *a*. Obviously, MCS and MCE differ only in the MUI scaling factor, and are identical when MUI is constant, confirming Dupuit's original insight. The advantage of MCE is that it is easily calculated when the indirect utility function and its derivatives are known, allows the introduction of policy change dimensions other than price, avoids the generally path-dependent definition of MCE, and usefully for retrospective analysis, expresses the change in well-being in units of the consumer's income in the "as is" scenario. The indirect utility function V has MUI constant, given p_0, if and only if it has an additively separable form $V(I,p_0,p_1) = \mu I/p_0 - G(p_1/p_0)$ for some function G and constant μ, in which case Roy's identity establishes that the demand for good 1, $D_1(I,p_0,p_1) = G'(p_1/p_0)/\mu$, is independent of income.

Dupuit's idea of solving the *inverse problem*, recovering utility from demand, was brought into mainstream economics at the end of the 19th century by William Stanley Jevons (1871), Francis Edgeworth (1881), Alfred Marshall (1890), Vilfredo Pareto (1906), and Eugen Slutsky (1915). MCS became the accepted measure of the change in consumer well-being. However, John Hicks (1939) observed that when the MUI = $\partial V(I,p_0,p_1)/\partial I$ is not constant, reducing income in scenario *b* by a transfer MCS will not necessarily leave the consumer indifferent between the scenarios. Hicks considered this a defect, and introduced two closely related alternative measures that correct it: *Hicksian Contingent Valuation* (HCV), the net decrease in scenario *b* income that equates utility in the two scenarios, and *Hicksian Equivalent Variation* (HEV), the net increase in scenario *a* income that equates utility in the two scenarios.

Importantly, the MCE, HCV, and HEV measures correspond to different consumer choice environments: The HCV measure assumes that the transfer is fulfilled in scenario *b* <u>before</u> the consumer makes a choice in that scenario, and the HEV measure assumes the transfer is fulfilled in scenario *a* <u>before</u> the scenario *a* choice. The MCE measure assumes that choices are made under actual market and income conditions in each scenario, <u>without</u> compensation, and that the post-choice transfer is determined <u>after</u> this as a remedy for the utility gain or loss from the change in scenario. Then, HCV is appropriate for prospective welfare analysis when the transfer is fulfilled before choice in scenario *b*, and HEV when the transfer is fulfilled before choice in scenario *a*. However, for retrospective welfare analysis where the objective is to redress past harm, or for prospective analysis where the transfers are hypothetical and not fulfilled, MCE is a more appropriate measure of what it takes to "make the consumer whole" following the choices the consumer did make or would have made in the uncompensated "as is" and "but for" scenarios. MCE is also appropriate for assessment of residual gains and losses subsequent to prospective analysis where an inexact compensation scheme is fulfilled.

HCV and HEV are often defined as areas to the left of income-compensated demand curves (i.e., demands with income adjusted as price changes to keep utility fixed at the scenario *a* or scenario *b* levels, respectively). However, their definition in terms of the indirect utility function, solutions to $V(I - HCV,p_0,p_{1b}) = V(I,p_0,p_{1a})$ and $V(I,p_0,p_{1b}) = V(I+HEV,p_0,p_{1a})$, are more revealing. Applying the mean value theorem, they satisfy

$$\text{(3)}\qquad \begin{aligned} \mathrm{HCV} &= [\mathrm{V}(I,\mathrm{p}_0,\mathrm{p}_{1\mathrm{b}}) - \mathrm{V}(I,\mathrm{p}_0,\mathrm{p}_{1\mathrm{a}})]/\mathrm{MUI}' \\ \mathrm{HEV} &= [\mathrm{V}(I,\mathrm{p}_0,\mathrm{p}_{1\mathrm{b}}) - \mathrm{V}(I,\mathrm{p}_0,\mathrm{p}_{1\mathrm{a}})]/\mathrm{MUI}'' \end{aligned},$$

where MUI' and MUI" are some intermediate values. Then, these two measures, the MCE measure from (2), and MCS are all proportional to the difference in utilities of the two scenarios, and differ only in scaling by the MUI valued at different points. Obviously, if the MUI is constant, then MCE, HCV, HEV, and MCS are identical, and in applications where the marginal utility of income varies little, they will be close approximations. MCE has a closed form when the indirect utility function is known, a computational advantage over HCV and HEV.

Samuelson (1947) and Hurwicz and Uzawa (1971) updated the Hicksian analysis using modern consumer theory, and their approach has been adapted to consumers making discrete choices by Diamond and McFadden (1974), Small and Rosen (1981), McFadden (1981, 1994, 1999, 2004, 2012, 2014), Yatchew (1985), and Zhou et al (2012). For the most part, this literature assumes that consumers are strictly neoclassical utility maximizers, with self-interest defined narrowly to include only personally purchased and consumed goods. Mostly, social motives are ignored and no allowance is made for ambiguities and uncertainties regarding tastes, budgets, hedonic attributes of goods and services, the reliability of transactions, or the consistency and completeness of preference maximization, and there is no distinction between the decision-utility postulated to determine market behavior and the experienced-utility of outcomes. Public and environmental goods are incorporated only if they have active margins that allow them to be valued from market behavior. The market demand functions of individual consumers are assumed to be completely observed, and consumers fully informed about policy regimes, so that utility can be recovered from the demand behavior it produces and the compensating transfers can be calculated and fulfilled exactly each consumer. The primary focus of welfare theory has been prospective, assuming that compensating transfers are fulfilled before consumer choices are made. The analysis has been fundamentally static, with the consumer pictured as making a once-and-for-all utility-maximizing choice for contingent deliveries of market goods, even if resolution of uncertainties and fulfillment of contracts extend over time; as in Debreu (1959). Analysis typically starts from prespecified scenarios, although in retrospective applications there are often substantive questions regarding the nature of the "but for" scenario, particularly when the "as is" scenario leads to experienced utility different from decision utility. Two further assumptions are tacit in most practical welfare calculations: First, policy scenario differences are limited in scope and magnitude, so that after accounting for a few major margins, general equilibrium effects can be neglected. Second, if compensating transfers are incomplete within a class of consumers, conducted say using a simple formula such as uniform transfers rather than an exact consumer-by-consumer calculation, the loss in social welfare from this imperfect redistribution can be neglected relative to the aggregate welfare change for the class.

We review these assumptions. Section 2 gives a foundation in consumer theory for the welfare calculus, with explicit treatment of discrete alternatives and their hedonic attributes. Section 3 restates the welfare measures in Section 1 for general applications, using the consumer theory of Section 2. Section 4 distinguishes retrospective and prospective policy applications of the welfare calculus. Section 5 discusses partial observability of individual consumer preferences, and its implications for welfare measurement and aggregation. Section 6 distinguishes decision-utility and experienced-utility foundations for calculation of well-being. Section 7 gives computational formulas for common policy problems. Section 8 contains an illustrative empirical application. Appendices collect relevant mathematical results on approximation, give properties of extreme-value distributed random variables, and give R-code for discrete welfare calculations.

2. CONSUMER FOUNDATIONS

A common starting assumption for welfare analysis is that consumers have "nice" demand functions that allow recovery of indirect utility. For example, Hurwicz and Uzawa (1971) give local and global sufficient conditions for recovery of money-metric indirect utility[2] when the market demand function is single-valued and smooth; see also Katzner (1970) and Border (2014). Another approach, originating in the revealed preference analysis of Samuelson (1948), Houthakker (1950), and Richter (1966), gives necessary and sufficient conditions for recovery of a preference order whose maximization yields the market demand function; Afriat (1967) and Varian (2006) provide constructive methods for recovery of utility under some conditions. Technical difficulties arise because quite strong smoothness and curvature conditions on utility are needed to assure smoothness properties on market demand, while preferences recovered from upper hemicontinuous demand functions are not necessarily continuous; see Peleg (1970), Rader (1973), Conniffe (2007). This section gives a restatement of the consumer theory behind welfare measurement, with extensions that include a "no local cliffs" Lipschitz continuity axiom on the preference map that avoids the Peleg-Rader problem and guarantees representation of preferences by utility, expenditure, indirect utility, and demand functions that satisfy (bi-)Lipschitz[3] conditions in economic variables. These results facilitate practical welfare measurement, and are of independent interest. Readers may find it useful to refer to Table 1 for notation, and consult as needed the technical material in the remainder of this section.

[2] Indirect utility is money-metric if the marginal utility of (real) income in a baseline scenario remains one as income changes.

[3] An increasing function is bi-Lipschitz if its left and right derivatives are bounded away from 0 and $+\infty$.

Table 1. Notation

m = *a*,*b*	"As-Is"/*baseline* policy/scenario *a* and "But-For"/*counterfactual* policy/scenario *b*
$s \in$ S	Finite-dimensional vector in a compact set S describing observed demographics and history of the decision-maker
$I_m \in [I^L, I^U]$	Consumer real income, in an interval $[I^L, I^U]$, with $0 < I^L < I^U < +\infty$, in scenario m
$j \in \mathbf{J}_m \subseteq \mathbf{J} = \{0,\ldots,J\}$	Mutually exclusive discrete choices (e.g., "products"), including "benchmark" or "no-purchase" alternatives that are not affected by policy change
$z_{jm} \in Z$	Vector $\mathbf{z}_m$ of observed hedonic attributes z_{jm} for alternatives $j \in \mathbf{J}_m$ in scenario m, in a compact finite-dimensional set Z
$\mathbf{q} \in Q' \subseteq Q$	Vector of the goods and services that are supplied in continuous quantities, in a finite rectangle $Q \equiv [0,\mathbf{q}^U]$ in n-dimensional Euclidean space, or in a subrectangle $Q' = [0,\mathbf{q}^A]$, where $\mathbf{q}^A$ is an upper bound on vectors that are affordable, $0 \ll \mathbf{q}^A \ll \mathbf{q}^U$
$w_{jm} = (\mathbf{q}, z_{jm}) \in W$	Consumption vector given discrete choice j in scenario m, in $W = Q\times Z$ or in $W' = Q'\times Z$
$p_{jm} \in P$	Real price, in a compact interval $P = [0,p^U]$ with $p^U > 0$, of discrete product j in scenario m; $\mathbf{p}_m$ is the vector with components p_{jm} for $j \in \mathbf{J}_m$
$\mathbf{r}_m \in R$	Finite-dimensional vector in a rectangle $R = [\mathbf{r}^L,\mathbf{r}^U]$, with $0 \ll \mathbf{r}^L \ll \mathbf{r}^U, \ll +\infty$, of real prices of the goods and services that are available in continuous quantities; benchmark $\mathbf{r}_a$
$\mathbf{r}_m \cdot \mathbf{q}_m + p_{jm} \le I_m$	Budget constraint given discrete alternative j in scenario m
$\succcurlyeq \in H$	A field H of complete transitive reflexive preference preorders $\succcurlyeq$ on $Q\times Z$, represented by sets $G(\succcurlyeq) \subseteq W\times W$ with $(w', w'') \in G(\succcurlyeq) \Leftrightarrow w' \succcurlyeq w''$
$U(\mathbf{q}, z, \succcurlyeq)$	A direct utility function conditioned on choice j in scenario m, defined on $Q'\times Z\times H$ as the minimum over $\mathbf{q}' \in Q^U$ of $\mathbf{r}_a\cdot\mathbf{q}'$ such that $(\mathbf{q}', z_{0a}) \succcurlyeq (\mathbf{q}, z_{jm}) \equiv w_{jm}$
$M(u, \mathbf{r}, z, \succcurlyeq)$	An expenditure function, the minimum over $\mathbf{q} \in Q$ of $\mathbf{r}\cdot\mathbf{q}$ such that $U(\mathbf{q}, z, \succcurlyeq) \ge u$
$\widetilde{V}(I, \mathbf{r}, z, \succcurlyeq)$	$\widetilde{V} \equiv I + \tilde{v}(I, \mathbf{r}, z, \succcurlyeq)$, a money-metric indirect utility function, the maximum of $U(\mathbf{q}, z, \succcurlyeq)$ subject to the budget constraint $\mathbf{r}\cdot\mathbf{q} \le I$
$\mathcal{V}(I_m, \mathbf{p}_m, \mathbf{r}_m, \mathbf{z}_m, \succcurlyeq)$	$\mathcal{V} = \max_{j\in\mathbf{J}_m} \widetilde{V}(I_m - p_{jm}, \mathbf{r}_m, z_{jm}, \succcurlyeq)$ unconditional maximum utility in scenario m
$\widetilde{P}_k(I, \mathbf{p}_m, \mathbf{r}_m, \mathbf{z}_m, s)$	The probability that choice k in scenario m attains maximum utility $\mathcal{V}$
$x_{jm} = X(I - p_{jm}, \mathbf{r}_m, z_{jm})$	A finite-dimensional vector of predetermined functions
$v(I_m - p_{jm}, \mathbf{r}, z_{jm}, \succcurlyeq)$	$v = x_{jm}\beta$ with $I = I_m$, parameters $\beta = \beta(\succcurlyeq)$, an approximation to $\tilde{v}\left(I_m - p_{jm}, \mathbf{r}, z_{jm}, \succcurlyeq\right)$
$V(I_m - p_{jm}, \mathbf{r}, z_{jm}, \succcurlyeq)$	$V = I_m - p_{jm} + x_{jm}\beta + \sigma\varepsilon_j$ approximation to $\widetilde{V}$ with additive EV1 "noise", $\sigma = \sigma(\succcurlyeq)$
$P_k(I, \mathbf{p}_m, \mathbf{r}, \mathbf{z}_m, s)$	$P_{km} = \mathbf{E}_{\beta\|s} \exp(\frac{x_{km}\beta - p_{km}}{\sigma}) / \sum_{j=0}^{J_m} \exp(\frac{x_{jm}\beta - p_{jm}}{\sigma})$ MMNL approximation to $\widetilde{P}_k(I, \mathbf{p}_m, \mathbf{r}, \mathbf{z}_m, s)$

Suppose consumers face scenarios m = a,b, and a universe of possible discrete alternatives indexed by a finite set $\mathbf{J} \equiv \{0,\ldots,J\}$. Let $\mathbf{J}_m \subseteq \mathbf{J}$ denote the set of alternatives available in the market under policy m, with $|\mathbf{J}_m|$ elements, and characterize then by *real* prices p_{jm} in a compact interval $P = [0,p^U]$ with $p^U > 0$, and *observed* hedonic attributes z_{jm} in a compact finite-dimensional set Z. Let $\mathbf{p}_m$ and $\mathbf{z}_m$ denote the vectors of p_{jm} and z_{jm} for $j \in \mathbf{J}_m$. Assume that there are alternatives that are always available and are unaffected by policy, including ordinarily "no purchase" alternatives that by convention are assigned zero price and attributes. Market goods supplied in continuous quantities are described by commodity vectors $\mathbf{q} \subseteq Q = [0,\mathbf{q}^U]$ with $0 \ll \mathbf{q}^U$, a bounded rectangle in n-dimensional space, with *real* market prices $\mathbf{r} \in R = [\mathbf{r}^L,\mathbf{r}^U]$, a commensurate bounded rectangle with $0 \ll \mathbf{r}^L \ll \mathbf{r}^U$. We assume that Z is a finite union of disjoint rectangles; this avoids technical complications and covers applications where measured attributes either vary continuously in some interval or take on a finite number of discrete levels.

Assume that consumers are characterized by a vector s of observed demographics and history, and by *real* income I_m in a bounded interval $[I^L,I^U]$. In many applications, $I_a = I_b$, but if changing from scenario a to scenario b entails an allocated net production cost or fulfilled transfer assessed as a lump sum net tax, then I_a and I_b will differ by the net amount. A consumer's market opportunities under policy m are summarized in prices $\mathbf{r}_m \in R$, and for $j \in \mathbf{J}_m$, attributes $z_{jm} \in Z$ and prices $p_{jm} \in P$, giving the budget constraint $\mathbf{r}_m \cdot \mathbf{q} \le I_m - p_{jm}$ for vectors $\mathbf{q} \in Q$ when a product j from $\mathbf{J}_m$ is chosen. Let $\mathbf{q}^A \in Q$ denote a vector that bounds all affordable vectors (i.e., $\mathbf{r}^L \cdot \mathbf{q} \le I^U$ implies $\mathbf{q} \ll \mathbf{q}^A$) and define $Q' = [0,\mathbf{q}^A]$. Let $\mathbf{z}$ denote the vector of attributes and $\mathbf{p}$ the vector of prices for the discrete alternatives in $\mathbf{J}$, and let $\mathbf{z}_m$ and $\mathbf{p}_m$ denote their subvectors for the available alternatives in $\mathbf{J}_m$.We adopt a description of consumers that is sufficiently flexible to encompass neoclassical preference maximization and some behavioral deviations, and can be made empirically tractable. Assume that consumers have complete transitive reflexive preference preorders $\succcurlyeq$ over vectors $(\mathbf{q},z) \in Q \times Z \equiv W$, that these preorders are predetermined and invariant with respect to current market opportunities, and that consumers are preference maximizers. Later, we consider the implications for identification of preferences and the welfare calculus when these neoclassical assumptions are relaxed. A preference preorder $\succcurlyeq$ is described by the non-empty set of pairs $((\mathbf{q}',z'),(\mathbf{q}'',z'')) \in W \times W$ that satisfy $(\mathbf{q}',z') \succcurlyeq (\mathbf{q}'',z'')$. Let $H \subseteq 2^{W \times W}$ denote the field of preference preorders of consumers in the population. We will assume that preferences for continuous goods are monotonic (i.e., $\mathbf{q}' \ge \mathbf{q}'' \Longrightarrow \mathbf{q}' \succcurlyeq \mathbf{q}''$), and that $\mathbf{q}^U$ is sufficiently large and continuous goods are sufficiently desirable so that they can substitute for any affordable $(\mathbf{q},z)$; i.e., $(\mathbf{q}^U,z_{0a}) \succcurlyeq (\mathbf{q}^A,z)$ for all $z \in Z$ and $\succcurlyeq \in H$.. We use the notation "$\succ$" for strict preference and the notation "$\sim$" for indifference. We use the Euclidean norm on Q, R, and Z; e.g., $\|\mathbf{q}\| = \sqrt{\mathbf{q} \cdot \mathbf{q}}$ for $\mathbf{q} \in Q$.

For non-empty subsets A,B of the metric space W×W, define the Hausdorff distance $h(A,B)$ to be the greatest lower bound of positive scalars η such that each set is contained in an η-neighborhood of the other; i.e., if $N_\eta(A)$

denotes the union of the open balls of radius η centered at the points in A, then h(A,B) is the greatest lower bound of η satisfying $B \subseteq N_\eta(A)$ and $A \subseteq N_\eta(B)$. The set W×W is compact, so h is bounded, and if A,B ∈ W×W are closed, then h(A,B) = 0 if and only if A = B. If the sets in H are all closed, then h is a metric on H termed the Hausdorff set metric, and H is precompact in its metric topology. We make a series of assumptions on preferences and budgets, beginning with a basic assumption on continuity of preferences:

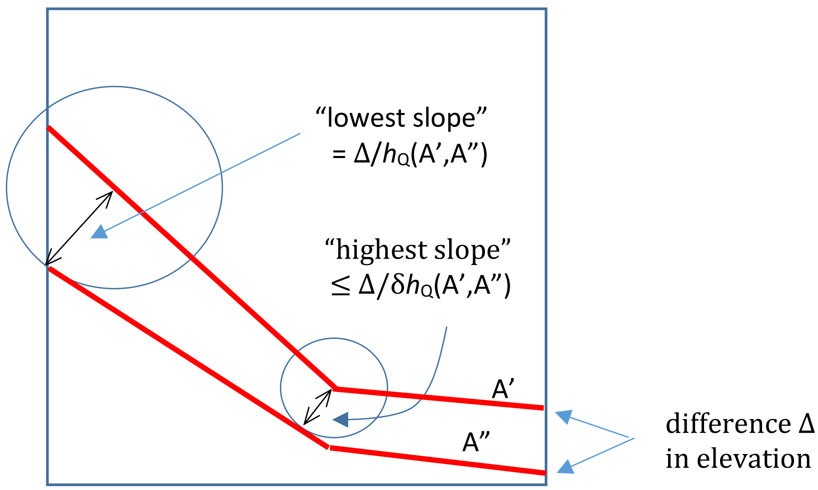
A1. If a sequence of preorders $\succcurlyeq^i \in H_+$ and sequences of consumption vectors $(w'^i,w''^i) \in G(\succcurlyeq^i)$ satisfy $h(G(\succcurlyeq^i),G(\succcurlyeq^0)) \to 0$, $w'^i \to w'^0$, and $w''^i \to w''^0$, then $G(\succcurlyeq^0) \in H$ and $(w'^0,w''^0) \in G(\succcurlyeq^0)$.

Since our attention is primarily on discrete choice, we will make strong and simple assumptions on continuous good preferences. Fix baseline values $(\mathbf{r}_a,z_a) \in R\times Z$. For $(\mathbf{q},z) \in Q\times Z$ and $\succcurlyeq \in H$, define $A(\mathbf{q},z,\succcurlyeq) = \{\mathbf{q}'\in Q \,|\, (\mathbf{q}',z_a) \succcurlyeq (\mathbf{q},z)\}$, the set of continuous commodity vectors $\mathbf{q}' \in Q$ that combined with "benchmark" attributes z_a are *at least as good a*s $(\mathbf{q},z)$. We will assume for $\mathbf{q} \in Q'$ that $\mathbf{q}^U \in A(\mathbf{q},z,\succcurlyeq)$, so this set is non-empty. Assumption A1 implies that $A(\mathbf{q},z,\succcurlyeq)$ is compact, and if $\mathbf{q}',\mathbf{q}'' \in Q'$, $z',z'' \in Z$, and $(\mathbf{q}',z') \succ (\mathbf{q}'',z'')$, then $A(\mathbf{q}',z',\succcurlyeq)$ is contained in the interior of $A(\mathbf{q}'',z'',\succcurlyeq)$. Assumption A2 strengthens our monotonicity requirement for continuous goods and imposes Lipschitz continuity conditions on preferences. Let $h_Q(A',A'')$ denote the Hausdorff distance between non-empty subsets $A',A'' \subseteq Q$. If $A' \subseteq A''$, then $h_Q(A',A'') \equiv \inf\{\eta > 0 | A'' \subseteq N_\eta(A')\}$. Assumptions A1 and A2 do not impose any convexity condition on preferences, but do require that the open quadrant to the northeast of any point in $A(\mathbf{q},z,\succcurlyeq)$ is contained in the interior of this set.

A2. For $\mathbf{q}',\mathbf{q}'' \in Q$, $z',z'' \in Z$, and $\succcurlyeq',\succcurlyeq'' \in H$, $\mathbf{q}'' \ll \mathbf{q}'$ implies $(\mathbf{q}',z') \succ' (\mathbf{q}'',z')$. If $\mathbf{q} \in Q'$, then $(\mathbf{q}^U,z_a) \succcurlyeq' (\mathbf{q},z')$. There exist scalars α, $\delta > 0$ such that (i) $(\mathbf{q}',z') \sim' (\mathbf{q}'',z'')$ implies $h_Q(A(\mathbf{q}',z',\succcurlyeq'),A(\mathbf{q}'',z'',\succcurlyeq'')) \le \alpha\cdot h(\succcurlyeq',\succcurlyeq'')$, (ii) $h_Q(A(\mathbf{q}',z',\succcurlyeq'),A(\mathbf{q}'',z'',\succcurlyeq'')) \le \alpha\cdot(|q' - q''| + |z' - z''| + h(\succcurlyeq',\succcurlyeq''))$, and (iii) $(\mathbf{q}',z') \succ' (\mathbf{q}'',z'')$ implies $\inf_{\mathbf{q}\in A(\mathbf{q}',z',\succcurlyeq')} \sup\{\eta > 0 \mid N_\eta(q) \subseteq A(\mathbf{q}'',z'',\succcurlyeq')\} \ge \delta h_Q(A(\mathbf{q}',z',\succcurlyeq'),A(\mathbf{q}'',z'',\succcurlyeq'))$.

For each $\succcurlyeq$, the $A(\mathbf{q},z,\succcurlyeq)$ are "at least as good as" sets whose boundaries define contours in Q, as illustrated in Figure 2. The Lipschitz continuity conditions (i) and (ii) rule out precipitous changes in $A(\mathbf{q},z,\succcurlyeq)$ when $(\mathbf{q},z,\succcurlyeq)$ changes. Let Δ denote the difference in elevation of the contours on some (arbitrary) scale. The line between boundary points in the upper contour sets $A' = A(\mathbf{q}',z',\succcurlyeq)$ and $A'' = A(\mathbf{q}'',z'',\succcurlyeq)$ that achieves their Hausdorff distance gives the lowest slope between them, $\Delta/h_Q(A',A'')$, while the highest slope is bounded by $\Delta/\delta h_Q(A',A'')$, where δ is the lower bound given by (iii). Then condition (iii) rules out "local cliffs" by bounding the ratio of the highest slope to the lowest slope as two contours converge. Preferences on a compact set that are representable by absolutely continuous utility functions with gradients that are bounded positive and finite satisfy the Lipschitz continuity condition. The utility function in $\mathbb{R}^2$ satisfying u = $q_1 - \sqrt{1-q_2}$ for $0 \le q_2 \le 1$ and u = $q_1 + \sqrt{q_2 - 1}$ for $q_2 > 1$ is an example that has a local cliff and fails to satisfy the condition at $q_2 = 1$.

Figure 2. Lowest and Highest Slopes between Contours



A last assumption confirms that income is in a range where consumer budgets are limiting but nevertheless allow choice of any available discrete alternative, and all affordable **q** are in Q':

A3. The consumption vector (0,*z*) is affordable for each $z \in$ Z (i.e., $I^L > p^U$), and **q** ∈ Q and $\mathbf{r}^L \cdot \mathbf{q} \le I^U \Longrightarrow \mathbf{q} \in$ Q'.

We next establish that A1-A3 are sufficient to guarantee well-behaved representations of preferences by utility, expenditure, and indirect utility functions.

Lemma 2.1. If A1, then each G(≽) ∈ H is a compact set, and H is a compact metric space with metric *h*.

Proof: For fixed ≽, A1 establishes that G(≽) is a non-empty closed subset of the compact space W×W. The properties of *h* and H are given in Aliprantis and Border (2006, Sections 3.16-3.18), particularly Definition 3.70, Theorem 3.85, and Corollary 3.95. ■

Lemma 2.2. Suppose A1 and A2, (**q**',*z*'), (**q**",*z*") ∈ Q'×Z, ≽ ∈ H, and (**q**',*z*') ≻ (**q**",*z*"). Let δ denote the bound from A2, **1** denote a vector of ones, and $\gamma = h_Q$(A(**q**', *z*',≽),A(**q**", *z*",≽)). If **q**'* is in the boundary of A(**q**',*z*',≽), **q**"* is in the boundary of A(**q**",*z*",≽), and **r** ∈ R, then **q**'* − γδ**r**/‖**r**‖ ∈ A(**q**",*z*",≽) and **q**"* + γ**1** ∈ A(**q**',*z*',≽).

Proof: By A2, all points in a neighborhood of $\mathbf{q}'^*$ with radius $\gamma\delta$ are contained in $A(\mathbf{q}'',z'',\succcurlyeq)$, giving the first result. If $\mathbf{q}''^* + \gamma\mathbf{1} \notin A(\mathbf{q}',z',\succcurlyeq)$, then no point within radius γ of $\mathbf{q}''^*$ is in $A(\mathbf{q}',z',\succcurlyeq)$, contradicting the definition of γ as the Hausdorff distance. ■

The next result uses the expenditure at baseline continuous good prices needed to achieve the level of satisfaction of a given vector of goods to define a well-behaved (i.e., bi-Lipschitz) utility function.

Theorem 2.3. Suppose A1 and A2. For each $(\mathbf{q},z) \in Q'\times Z$ and $\succcurlyeq \in H$, define

(5) $U(\mathbf{q},z,\succcurlyeq) = \min\{\mathbf{r}_a\cdot\mathbf{q}' \mid \mathbf{q}' \in A(\mathbf{q},z,\succcurlyeq)\}$.

Then, U is a continuous *direct utility function* on $Q'\times Z\times H$; i.e., U is continuous in its arguments and $U(\mathbf{q}',z',\succcurlyeq) \geq U(\mathbf{q}'',z'',\succcurlyeq) \Leftrightarrow (\mathbf{q}',z') \succcurlyeq (\mathbf{q}'',z'')$. Further, U is locally non-satiated with a range $U(Q',z,\succcurlyeq)$ contained in the bounded interval $[0,\mathbf{r}_a\cdot\mathbf{q}^U]$, is Lipschitz in $(\mathbf{q},z,\succcurlyeq)$, and is *bi-Lipschitz* in $A = A(q,z,\succcurlyeq)$; i.e., there exist scalars $\alpha_U, \lambda_U > 0$ such that $\mathbf{q}',\mathbf{q}'' \in Q'$ and $z',z'' \in Z$ imply

(6)
$$h_Q(A(\mathbf{q}',z',\succcurlyeq),A(\mathbf{q}'',z'',\succcurlyeq))\cdot\lambda_U \geq |U(\mathbf{q}',z',\succcurlyeq) - U(\mathbf{q}'',z'',\succcurlyeq)| \geq h_Q(A(\mathbf{q}',z',\succcurlyeq),A(\mathbf{q}'',z'',\succcurlyeq))/\lambda_U$$
$$|U(\mathbf{q}',z',\succcurlyeq') - U(\mathbf{q}'',z'',\succcurlyeq'')| \leq \alpha_U\cdot(|\mathbf{q}'-\mathbf{q}''| + |z'-z''| + h(\succcurlyeq',\succcurlyeq'')) .$$

Proof: A1 and A2 imply that $A(\mathbf{q},z,\succcurlyeq)$ is non-empty and closed, hence compact, so (5) is well-defined. Suppose $(\mathbf{q}',z') \succcurlyeq (\mathbf{q}'',z'')$. If $(\mathbf{q},z_a) \in A(\mathbf{q}',z',\succcurlyeq)$, then $(\mathbf{q},z_a) \succcurlyeq (\mathbf{q}',z')$ and transitivity implies $(\mathbf{q},z_a) \succcurlyeq (\mathbf{q}'',z'')$, and hence $(\mathbf{q},z_a) \in A(\mathbf{q}'',z'',\succcurlyeq)$. Therefore, $\mathbf{r}_a\cdot\mathbf{q} \geq U(\mathbf{q}'',z'',\succcurlyeq)$, and hence $U(\mathbf{q}',z',\succcurlyeq) \geq U(\mathbf{q}'',z'',\succcurlyeq)$. A1 implies that U is continuous in its arguments. In particular, U is sequentially continuous in $\succcurlyeq$ in the Hausdorff metric topology on H, or equivalently the topology of closed convergence.

If $\mathbf{q}^{*\prime} \in \operatorname{argmin}\{\mathbf{r}_a\cdot\mathbf{q} \mid \mathbf{q} \in A(\mathbf{q}',z',\succcurlyeq)\}$ and $\mathbf{q}^{*\prime\prime} \in \operatorname{argmin}\{\mathbf{r}_a\cdot\mathbf{q} \mid \mathbf{q} \in A(\mathbf{q}'', z'',\succcurlyeq)\}$, then $U(\mathbf{q}', z',\succcurlyeq) - U(\mathbf{q}'', z'',\succcurlyeq) = \mathbf{r}_a\cdot(\mathbf{q}^{*\prime} - \mathbf{q}^{*\prime\prime})$. Let $\gamma \equiv h_Q(A(\mathbf{q}',z',\succcurlyeq),A(\mathbf{q}'',z'',\succcurlyeq))$. Lemma 2.2 with $\mathbf{r} = \mathbf{r}_a$ implies $\tilde{\mathbf{q}}'' = \mathbf{q}^{*\prime} - \delta\gamma\mathbf{r}_a/\|\mathbf{r}_a\| \in A(\mathbf{q}'',z'',\succcurlyeq)$ and $\tilde{\mathbf{q}}' = \mathbf{q}^{*\prime\prime} + \gamma\mathbf{1} \in A(\mathbf{q}',z',\succcurlyeq)$. Then $\mathbf{r}_a\cdot(\mathbf{q}^{*\prime} - \tilde{\mathbf{q}}'') = \delta\gamma\|\mathbf{r}_a\| \leq \mathbf{r}_a\cdot(\mathbf{q}^{*\prime} - \mathbf{q}^{*\prime\prime})$ and $\mathbf{r}_a\cdot(\tilde{\mathbf{q}}' - \mathbf{q}^{*\prime\prime}) = \gamma\mathbf{r}_a\cdot\mathbf{1} \geq \mathbf{r}_a\cdot(\mathbf{q}^{*\prime} - \mathbf{q}^{*\prime\prime})$. Defining $\lambda_U = \max(\mathbf{r}_a\cdot\mathbf{1},1/\delta\|\mathbf{r}_a\|)$, this gives the first row of (6). Then, the first Lipschitz condition in A2 gives the second inequality in (6),

$$|U(\mathbf{q}',z',\succcurlyeq) - U(\mathbf{q}'',z'',\succcurlyeq)| \leq \lambda_U h_Q(A(\mathbf{q}',z',\succcurlyeq),A(\mathbf{q}'',z'',\succcurlyeq)) \leq \lambda_U\alpha\cdot(|q'-q''| + |z'-z''| + h(\succcurlyeq',\succcurlyeq'')).$$ ■

Theorem 2.4. Suppose A1 and A2, and $U:Q'\times Z\times H \to [0,\mathbf{r}_a\cdot\mathbf{q}^U]$ from (5). For each $(\mathbf{r},z) \in R\times Z$, $\succcurlyeq\, \in H$, and $u \in U(Q',z,\succcurlyeq)$, define

(7) $M(u,\mathbf{r},z,\succcurlyeq) = \min_{\mathbf{q}'\in Q'}\{\mathbf{r}\cdot\mathbf{q}' \mid U(\mathbf{q}',z,\succcurlyeq) \geq u\}$.

Then M is an *expenditure function* that is continuous in its arguments, and concave, linear homogeneous, and non-decreasing in $\mathbf{r}$. Further, M satisfies $M(u,\mathbf{r}_a,z_a,\succcurlyeq) \equiv u$, is Lipschitz in $(\mathbf{r},z,\succcurlyeq)$, and is *bi-Lipschitz and increasing* in u; i.e., there exists scalars α_M, $\lambda_M > 0$ such that $(\mathbf{r}',z',\succcurlyeq')$, $(\mathbf{r}'',z'',\succcurlyeq'') \in R\times Z\times H$ and $u',u'' \in U(Q',z,\succcurlyeq)$ with $u' \geq u''$ imply

$$(8)\qquad \begin{array}{c} (u' - u'')\lambda_M \;\geq\; M(u',\mathbf{r}',z',\succcurlyeq') - M(u'',\mathbf{r}',z',\succcurlyeq') \;\geq\; (u' - u'')/\lambda_M \\ |M(u',\mathbf{r}',z',\succcurlyeq') - M(u',\mathbf{r}'',z'',\succcurlyeq'')| \leq \alpha_M\cdot(|r' - r''| + |z' - z''| + h(\succcurlyeq',\succcurlyeq'')) \end{array}.$$

Proof: Mas-Colell, Whinston, and Green (1995, Proposition 3E2) demonstrate that M is concave, linear homogeneous, and non-decreasing in $\mathbf{r}$. The continuity of M in its arguments follows from the Berge Maximum Theorem (Aliprantis and Border, 2006, Theorem 17.31). Suppose $(\mathbf{r}',z',\succcurlyeq')$, $(\mathbf{r}'',z'',\succcurlyeq'') \in R\times Z\times H$, and $u' \geq u''$. Consider $\mathbf{q}^{*\prime} \in \operatorname{argmin}\{\mathbf{r}'\cdot\mathbf{q} \mid U(\mathbf{q},z',\succcurlyeq') \geq u'\}$ and $\mathbf{q}^{*\prime\prime} \in \operatorname{argmin}\{\mathbf{r}'\cdot\mathbf{q} \mid U(\mathbf{q},z',\succcurlyeq') \geq u''\}$. Let $\gamma = h_Q(A(\mathbf{q}^{*\prime},z',\succcurlyeq),A(\mathbf{q}^{*\prime\prime},z',\succcurlyeq))$. From Lemma 2.2, $\tilde{\mathbf{q}}'' = \mathbf{q}^{*\prime} - \delta\gamma\mathbf{r}/\|\mathbf{r}\| \in A(\mathbf{q}^{*\prime\prime},z',\succcurlyeq')$ and $\tilde{q}' = \mathbf{q}^{*\prime\prime} + \gamma\mathbf{1} \in A(\mathbf{q}^{*\prime},z',\succcurlyeq')$, so that $\mathbf{r}\cdot(\mathbf{q}^{*\prime} - \tilde{\mathbf{q}}'') = \delta\gamma\|\mathbf{r}\| \leq \mathbf{r}\cdot(\mathbf{q}^{*\prime} - \mathbf{q}^{*\prime\prime}) = M(u',r',z',\succcurlyeq') - M(u'',r',z',\succcurlyeq') \leq \mathbf{r}\cdot(\tilde{\mathbf{q}}' - \mathbf{q}^{*\prime\prime}) = \gamma\mathbf{r}\cdot\mathbf{1}$. This establishes the first row inequality in (8) with $\lambda_M = \lambda_U \max(\mathbf{r}^U\cdot\mathbf{1}, 1/\delta\|\mathbf{r}^L\|)$.

Next consider $\mathbf{q}^{*\prime} \in \underset{U(\mathbf{q},z',\succcurlyeq')\geq u'}{\operatorname{argmin}} \mathbf{r}'\cdot\mathbf{q}$ and $\mathbf{q}^{*\prime\prime} \in \underset{U(\mathbf{q},z'',\succcurlyeq'')\geq u'}{\operatorname{argmin}} \mathbf{r}''\cdot\mathbf{q}$, and define $\gamma = h_Q(A(\mathbf{q}^{*\prime\prime},z'',\succcurlyeq''),A(\mathbf{q}^{*\prime\prime},z',\succcurlyeq'))$ $\leq \alpha(|z' - z''| + h(\succcurlyeq',\succcurlyeq''))$. Then, $A(q^{*\prime},z',\succcurlyeq') = A(q^{*\prime\prime},r'',\succcurlyeq')$. If $M(u',\mathbf{r}',z',\succcurlyeq') > M(u',\mathbf{r}'',z'',\succcurlyeq')$, then $q^{*\prime\prime} + \gamma\mathbf{1} \in A(\mathbf{q}^{*\prime\prime},z',\succcurlyeq')$, implying $M(u',\mathbf{r}',z',\succcurlyeq') - M(u',\mathbf{r}'',z'',\succcurlyeq') \leq \gamma\mathbf{r}'\cdot\mathbf{1} + |(\mathbf{r}'' - \mathbf{r}')\cdot\mathbf{q}^{*\prime\prime}| \leq \|\mathbf{r}^U\|\alpha(|z' - z''| + h(\succcurlyeq',\succcurlyeq''))$

$\leq |\mathbf{r}'' - \mathbf{r}'|\cdot\|\mathbf{q}^U\| + \|\mathbf{r}^U\|\alpha(|z' - z''| + h(\succcurlyeq',\succcurlyeq''))$, proving the second row in (8) with $\alpha_M = \max(\|\mathbf{q}^U\|, \|\mathbf{r}^U\|\alpha)$. ■

Theorem 2.5. Suppose A1-A3 and $U:Q'\times Z\times H \to [0,\mathbf{r}_a\cdot\mathbf{q}^U]$ from (5). For $I \in [I^L - p^U, I^U]$, $\mathbf{r} \in R$, $z \in Z$, and $\succcurlyeq\, \in H$, define *a money-metric indirect utility function*

(9) $V(I,\mathbf{r},z,\succcurlyeq) \equiv \max_{\mathbf{q}\in Q'}\{U(\mathbf{q},z,\succcurlyeq) \mid \mathbf{r}\cdot\mathbf{q} \leq I\}$

satisfying $V(I,\mathbf{r}_a,z_a,\succcurlyeq) \equiv I$ that is continuous in its arguments with a range contained in the bounded interval $[0,\mathbf{r}_a\cdot\mathbf{q}^U]$, is quasi-convex and homogeneous of degree zero in $(I,\mathbf{r})$, is non-increasing in $\mathbf{r}$, is Lipschitz in $(\mathbf{r},z,\succcurlyeq)$, and is bi-Lipschitz increasing in I; i.e., there exist scalars $\alpha_V \geq \mathbf{r}_a \cdot \mathbf{q}^U$ and $\lambda_V > 0$ such that $I' > I''$, $\mathbf{r}',\mathbf{r}'' \in R$, $z',z'' \in Z$, and $\succcurlyeq',\succcurlyeq'' \in H$ imply

$$(10)\quad \begin{array}{c} \lambda_V(I' - I'') \;\geq\; V(I',\mathbf{r}',z',\succcurlyeq') - V(I'',\mathbf{r}',z',\succcurlyeq') \;\geq\; (I' - I'')/\lambda_V \\ |V(I',\mathbf{r}',z',\succcurlyeq') - V(I'',\mathbf{r}'',z'',\succcurlyeq'')| \leq \alpha_V(|I' - I''| + |\mathbf{r}' - \mathbf{r}| + |z'-z| + h(\succcurlyeq',\succcurlyeq'')) \end{array}.$$

Further, $V(I,\mathbf{r},z,\succcurlyeq)$ is twice continuously differentiable in $(I,\mathbf{r})$ except on a set of measure zero, and continuous good demands are almost everywhere in $(I,\mathbf{r})$ single-valued and continuously differentiable, and satisfy Roy's identity, $q = D(I, \mathbf{r}, z, \succcurlyeq) \equiv -\frac{\partial V(I,\mathbf{r},z,\succcurlyeq)/\partial \mathbf{r}}{\partial V(I,\mathbf{r},z,\succcurlyeq)/\partial I}$.

Proof: Theorem 2.3 implies that U is continuous in its arguments, and A3 assures that the budget set for **q** is a non-empty subset of Q' for all discrete choices. Then, (9) is well-defined. Local non-satiation from A2 implies that V is the inverse with respect to u of $I = M(u,\mathbf{r},z,\succcurlyeq)$; see Mas-Colell, Whinston, and Green (1995, Propositions 3D3 and 3E1). Result (8) then implies the first row of (10) with $\lambda_V = \lambda_U$. The continuity of M in its arguments from Lemma 2 and the bi-Lipschitz condition (8) imply V is continuous in its arguments and Lipschitz-continuous in $(\mathbf{r},z,\succcurlyeq)$. It is immediate from the properties of M that V is homogeneous of degree zero in $(I,\mathbf{r})$, and non-increasing in $\mathbf{r}$. Consider budgets $(I^i,\mathbf{r}^i)$ for i = 0,1 and $(I^\theta,\mathbf{r}^\theta) = \theta(I^0,\mathbf{r}^0) + (1-\theta)(I^1,\mathbf{r}^1)$ for $\theta \in (0,1)$, and let q^* denote a maximand of (5) subject to the budget $(I^\theta,\mathbf{r}^\theta)$. Then $V(I^\theta,\mathbf{r}^\theta,z,\succcurlyeq) = U(q^*,z,\succcurlyeq)$. But $\mathbf{r}^\theta \cdot q^* \leq I^\theta$ implies either or both $\mathbf{r}^0 \cdot q^* \leq I^0$ or $\mathbf{r}^1 \cdot q^* \leq I^1$, and therefore either $V(I^0,\mathbf{r}^0,z,\succcurlyeq) = U(q^*,z,\succcurlyeq)$ or $V(I^1,\mathbf{r}^1,z,\succcurlyeq) = U(q^*,z,\succcurlyeq)$, so that V is quasi-convex: $V(I^\theta,\mathbf{r}^\theta,z,\succcurlyeq) \leq \max\{V(I^0,\mathbf{r}^0,z,\succcurlyeq), V(I^1,\mathbf{r}^1,z,\succcurlyeq)\}$. From the definition of quasi-convexity, there exists an increasing transformation ψ such that such that $v(I,\mathbf{r},z,\succcurlyeq) = \psi(V(I,\mathbf{r},z,\succcurlyeq))$ is a convex function of (I,r). Results of Rademacher (1919) and Alexandrov (1939) establish that since $v(I,\mathbf{r}_a,z_a,\succcurlyeq) \equiv \psi(I)$ is convex in I, ψ is bi-Lipschitz on $(I^L - p^U, I^U)$, continuously differentiable in I except possibly on a countable set, and almost everywhere twice continuously differentiable in I. Then, $\psi^{-1}(v)$ is also bi-Lipschitz and increasing, and hence continuously differentiable except on a countable set. This implies that $V(I,\mathbf{r},z,\succcurlyeq) = \psi^{-1}(v(I,\mathbf{r},z,\succcurlyeq))$ is increasing and bi-Lipschitz in I, and hence continuously differentiable in I except for a countable set, and almost everywhere twice continuously differentiable in $(I,\mathbf{r})$. Then the Roy (1947) identity applied to $v(I,\mathbf{r}_a,z,\succcurlyeq)$, or equivalently to $V(I,\mathbf{r},z,\succcurlyeq)$, establishes that continuous good demands are almost surely single-valued and continuously differentiable in $(I,\mathbf{r})$. ■

Lemma 2.6. Suppose A1-A3, the direct utility function $U(\mathbf{q},z,\succcurlyeq)$ from (5), and its associated money metric indirect utility function $V(I,\mathbf{r},z,\succcurlyeq)$ from (9). For $\mathbf{q} \in Q'$, define $U^*(\mathbf{q},z,\succcurlyeq) = \min_{\mathbf{r}\in R} V(\mathbf{r}\cdot\mathbf{q},\mathbf{r},z,\succcurlyeq)$. Then $U^*(\mathbf{q},z,\succcurlyeq)$ is quasi-concave and R-monotone[4], $U^*(\mathbf{q},z,\succcurlyeq) \ge U(\mathbf{q},z,\succcurlyeq)$, with equality if for some $\mathbf{r} \in R$, the conditions $\mathbf{q}' \in Q'$ and $\mathbf{r}\cdot\mathbf{q}' \le \mathbf{r}\cdot\mathbf{q}$ imply $U(\mathbf{q}',z,\succcurlyeq) \le U(\mathbf{q},z,\succcurlyeq)$. Then, U and U* are observationally equivalent; i.e., $V(I,\mathbf{r},z,\succcurlyeq) \equiv \max_{\mathbf{q}\in Q'}\{U^*(\mathbf{q},z,\succcurlyeq) \mid \mathbf{r}\cdot\mathbf{q} \le I\}$ and the continuous good demands from maximization of U and U* subject to the budget constraint $\mathbf{r}\cdot\mathbf{q} \le I$ coincide except on a set of $(I,\mathbf{r})$ of measure zero.

Proof: $V(\mathbf{r}\cdot\mathbf{q},\mathbf{r},z,\succcurlyeq) = \max\{U(\mathbf{q}',z,\succcurlyeq) \mid \mathbf{r}\cdot\mathbf{q}' \le \mathbf{r}\cdot\mathbf{q}\} \ge U(\mathbf{q},z,\succcurlyeq)\}$ implies $U^*(\mathbf{q},z,\succcurlyeq) \ge U(\mathbf{q},z,\succcurlyeq)$. Suppose for some $\mathbf{r} \in R$, the conditions $\mathbf{q}' \in Q'$ and $\mathbf{r}\cdot\mathbf{q}' \le \mathbf{r}\cdot\mathbf{q}$ imply $U(\mathbf{q}',z,\succcurlyeq) \le U(\mathbf{q},z,\succcurlyeq)$, so that $\mathbf{q}'$ is a maximand of U subject to this budget constraint. Then $U^*(\mathbf{q},z,\succcurlyeq) \le V(\mathbf{r}\cdot\mathbf{q},\mathbf{r},z,\succcurlyeq) \le U(\mathbf{q},z,\succcurlyeq)$. Hence, U and U* are observationally equivalent. Suppose $\mathbf{q}^0,\mathbf{q}^1 \in Q'$ and $\mathbf{q}^\theta = \theta\mathbf{q}^0 + (1-\theta)\mathbf{q}^1$ for $\theta \in (0,1)$, and let $\mathbf{r}^\theta \in R$ be such that $U^*(\mathbf{q}^\theta,z,\succcurlyeq) = V(\mathbf{r}^\theta\cdot\mathbf{q}^\theta,\mathbf{r}^\theta,z,\succcurlyeq)$. Since $\mathbf{r}^\theta\cdot\mathbf{q}^\theta \ge \min\{\mathbf{r}^\theta\cdot\mathbf{q}^0,\mathbf{r}^\theta\cdot\mathbf{q}^1\}$, $U^*(\mathbf{q}^\theta,z,\succcurlyeq) = V(\mathbf{r}^\theta\cdot\mathbf{q}^\theta,\mathbf{r}^\theta,z,\succcurlyeq) \ge \min\{V(\mathbf{r}^\theta\cdot\mathbf{q}^0,\mathbf{r}^\theta,z,\succcurlyeq), V(\mathbf{r}^\theta\cdot\mathbf{q}^1,\mathbf{r}^\theta,z,\succcurlyeq)\} \ge \min\{U^*(\mathbf{q}^0,z,\succcurlyeq), U^*(\mathbf{q}^1,z,\succcurlyeq)\}$, so U* is quasi-concave in $\mathbf{q}$. Suppose $\mathbf{r}\cdot\mathbf{q}'' > \mathbf{r}\cdot\mathbf{q}'$ for all $r \in R$. Then there exists $\mathbf{r}''$ satisfying $U^*(\mathbf{q}'',z,\succcurlyeq) = V(\mathbf{r}''\cdot\mathbf{q}'',\mathbf{r}'',z,\succcurlyeq) > V(\mathbf{r}''\cdot\mathbf{q}',\mathbf{r}'',z,\succcurlyeq) \ge U^*(\mathbf{q}',z,\succcurlyeq)$, so U* is R-monotone. ■

Let $V(I - p_{jm},\mathbf{r}_m,z_{jm},\succcurlyeq)$ denote the indirect utility function from (9) for discrete alternative j with attributes z_{jm}, price p_{jm}, and income $I - p_{jm}$ remaining for purchase of continuous goods. The consumer who chooses $j \in \mathbf{J}_m$ and $q \in Q$ to maximize utility subject to the budget constraint $\mathbf{r}_m\cdot\mathbf{q} + p_{jm} \le I$ then achieves *unconditional indirect utility*

$$(11)\qquad u = \mathcal{V}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq) \equiv \max_{j\in\mathbf{J}_m} V(I - p_{jm},\mathbf{r}_m,z_{jm},\succcurlyeq)\,.$$

Associated with (11) is an *unconditional expenditure function* $I = \mathcal{M}(u,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq)$ obtained as an implicit solution of (11), or equivalently as

$$(12)\qquad I = \mathcal{M}(u,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq) \equiv \min_{j\in J_m}[M(u,\mathbf{r}_m,z_{jm},\succcurlyeq) + p_{jm}]\,.$$

Theorems 2.4 and 2.5 imply that $\mathcal{M}$ is bi-Lipschitz increasing in u and $\mathcal{V}$ is bi-Lipschitz increasing in I.

Next characterize the choices and demands that achieve (11). For $k \in \mathbf{J}_m$, and $I \in [I^L,I^U]$, $\mathbf{p}_m \in P^{|J_m|}$, $\mathbf{z}_m \in Z^{|J_m|}$, and $\mathbf{r}_m \in R$, define the set of preferences that make alternative k uniquely optimal,

$$(13)\qquad H^k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m) = \{\succcurlyeq\in H \mid V(I - p_{km},\mathbf{r}_m,z_{km},\succcurlyeq) > V(I - p_{jm},\mathbf{r}_m,z_{jm},\succcurlyeq) \text{ for } j \in \mathbf{J}_m\backslash\{k\}\}$$

---

[4] A function $u(\mathbf{q})$ is quasi-concave if $0 < \theta < 1$ implies $u(\theta\mathbf{q}'+(1-\theta)\mathbf{q}'') \ge \min(u(\mathbf{q}'),u(q''))$, and is R-monotone if $\mathbf{r}\cdot\mathbf{q}' \ge \mathbf{r}\cdot\mathbf{q}''$ for all $\mathbf{r} \in R$ implies $u(\mathbf{q}') \ge u(\mathbf{q}'')$.

and let $H^{\#}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m) = \cup_{k\in\mathbf{J}_m} H^k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)$ denote the set of all preferences that result in a unique utility-maximizing choice. For $\succcurlyeq \in H^{\#}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)$, choice is indicated by

$$(14) \qquad \delta_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq) = \mathbf{1}_{H^k(I,\mathbf{p}_m,\mathbf{r},\mathbf{z}_m)}(\succcurlyeq) \equiv \begin{cases} 1 & \text{if } \succcurlyeq \in H^k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m) \\ 0 & \text{otherwise} \end{cases},$$

and using Theorem 2.5 and Roy's identity, continuous good demands are given for almost all (I,$\mathbf{r}$) by

$$(15) \qquad D(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq) = -\sum_{k\in\mathbf{J}_m} \delta_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq)\cdot\frac{\partial V(I-p_{km},\mathbf{r}_m,z_{km},\succcurlyeq)/\partial\mathbf{r}}{\partial V(I-p_{km},\mathbf{r}_m,z_{km},\succcurlyeq)/\partial I}.$$

In applications, the preferences $\succcurlyeq$ in (9) are unobserved and are heterogeneous across consumers. Limited observations on the market choices of a single consumer allow only partial identification of preferences, insufficient to estimate (9) with precision. Therefore we treat preferences as a "random effect" in (9), with a probability $F^H(\cdot|s)$ on the field of preferences H whose salient features we can hope to identify from market data. Our final preference assumption is that preference heterogeneity is almost surely sufficient to break ties:

A4. $F^H(H^{\#}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)|s) \equiv 1$ for $(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m) \in [I^L,I^U]\times P^{J_m+1}\times R\times Z^{J_m+1}$ and $s \in S$.

This assumption can be restated as a probabilistic transversality condition that the distribution of the vector of indirect utilities for the various alternatives is of full dimension and absolutely continuous; see Shannon (2006). Given Assumption A4, the discrete alternatives $k \in \mathbf{J}_m$, are chosen with probabilities

$$(16) \qquad P_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,s) = F^H(H^k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)|s) \equiv \partial\mathbf{E}_{\succcurlyeq|s}\mathcal{V}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq)/\,\partial v_{jm},$$

with the last form of (16) coming from the interpretation of $\mathbf{E}_{\succcurlyeq|s}\mathcal{V}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq)$ as a *Choice-Probability-Generating-Function* (CPGF) with $v_{jm} \equiv V(I-p_{jm},\mathbf{r}_m,z_{jm},\succcurlyeq)$ for $j \in \mathbf{J}_m$ as arguments; see Fosgerau, McFadden, and Bierlaire (2013).[5] The conditional probability of continuous good demand in a measurable set $B \subseteq Q$, given choice k, is $F^H(\{\succcurlyeq \in H^k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m) \mid s, D(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq) \in B\})/P_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,s)$, and the conditional probability of $\succcurlyeq$ given choice k satisfies $F^H(A|s,k) = F^H(A\cap H^k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)|s)/P_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,s)$ for measurable $A \subseteq H$.

For welfare applications, the representation in (9) and (11) of a population preference field satisfying Assumptions A1-A4 has to be translated into a system that is practical for estimation and calculation. One approach is direct non-parametric estimation of $\mathbf{E}_{\succcurlyeq|s}\mathcal{V}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq)$ using the property (16) that its gradient

[5] When $\mathbf{E}_{\succcurlyeq|s}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq)$ is additively linear in income, the CPGF coincides with the *social surplus function* introduced by McFadden (1981). The greater generality of the CPGF comes from recognizing that treating the v_{jm} as linear perturbations of utility gives the gradient property even when real income is not linear and additive in the indirect utility function.

equals the vector of choice probabilities; see Bhattacharya (2015, 2017). This approach can be sharpened by using the Lipschitz properties of (11) and adapting the Hall and Yachew (2007) method for nonparametric estimation of a function and its derivatives. A limitation of a fully nonparametric approach is that its regularities are local, so that it has difficulty predicting consumer outcomes when policies require non-local extrapolation. A second approach to practical analysis is the method of sieves, utilizing a net of finite-parameter approximations to the consumer preference field. Advantages of this approach are that it requires at an entry level only the finite-parameter methods and software employed in traditional applied economics, and that it is relatively easy to impose structural restrictions that support plausible policy extrapolation. In this paper, we provide a foundation for this second approach by showing that the field of indirect utility functions (9) with the properties given by Assumptions A1-A4 can be approximated uniformly by a practical finitely-parameterized family, with random parameters in the population that have a finitely parameterized distribution. Then, this family can be estimated from observed choices in sufficiently rich arrays of market environments faced by samples of consumers, and the estimated family can be used to carry out welfare calculations with no essential loss of generality.

Theorem 2.7. Suppose A1-A4. Let $\widetilde{\mathrm{V}}(I,\mathbf{r}_m,z_{jm},\succcurlyeq)$ for $(I,\mathbf{r}_m,z_{jm},\succcurlyeq) \in [I^L-p^U,I^U]\times R\times Z\times H$ denote the true indirect utility function from Theorem 2.5, and define $\tilde{\mathrm{v}}\left(I,p_{\mathrm{jm}},\mathbf{r}_{\mathrm{m}},z_{\mathrm{jm}},\succcurlyeq\right) \equiv \widetilde{\mathrm{V}}(I-p_{jm},\mathbf{r}_m,z_{jm},\succcurlyeq) - I$ on $[I^L,I^U]\times[0,p^U]\times R\times Z\times H$. Given a small scalar $\gamma \in (0,1)$, there exists a bound $\eta = -\ln(\gamma/4|\mathbf{J}_m|)$; a vector of predetermined twice continuously differentiable functions $X:[I^L-p^U,I^U]\times R\times Z \longrightarrow \mathbb{R}^N$ drawn from a Schauder basis[6] for the space $\mathfrak{C}([I^L-p^U,I^U]\times R\times Z)$; a commensurate vector of Lipschitz-continuous real functions β from a compact subset $\mathcal{B} \subseteq \mathfrak{C}(\mathrm{H},\mathbb{R}^N)$, and a Lipschitz-continuous real function $\sigma:\mathrm{H} \longrightarrow [\sigma^L,\sigma^U]$ from a compact subset $\mathcal{S} \subseteq \mathfrak{C}(\mathrm{H},[\sigma^L,\sigma^U])$ with $\sigma^L > 0$ and $\sigma^U < \gamma/2\eta$; and independent standard type I extreme value (EV1) distributed random variables ε_j such that:

(i) There is an approximate indirect utility function[7]

$$(17)\qquad \mathrm{V}(I - p_{jm},\mathbf{r}_m,z_{jm},\beta,\sigma,\varepsilon_j) = I + \mathrm{v}(I,p_{jm},\mathbf{r}_m,z_{jm},\beta) + \sigma\varepsilon_j$$

[6] A Schauder basis may be polynomials, Fourier series, or other series of functions that span the space of continuous functions on a compact finite-dimensional space. The basis may be tailored to reduce the number of terms required to achieve a given tolerance.

[7] The approximation V is not guaranteed to satisfy the slope and curvature properties of $\widetilde{\mathrm{V}}$, but at each point where $\widetilde{\mathrm{V}}$ is twice continuously differentiable with non-zero slopes and a non-singular (bordered) hessian, the approximation V for a sufficiently small tolerance γ will also have these properties and preserve signs, and hence locally have the same slope and curvature properties as $\widetilde{\mathrm{V}}$.

on $[I^L - p^U, I^U] \times R \times Z \times \mathcal{B} \times \mathcal{S} \times \mathbb{R}$ with $v(I,p_{jm},\mathbf{r}_m,z_{jm},\beta) \equiv X(I - p_{jm}, \mathbf{r}_m, z_{jm}) \cdot \beta - p_{jm}$ such that $|\tilde{v}(I,p_{jm},\mathbf{r}_m,z_{jm},\succcurlyeq) - v(I,p_{jm},\mathbf{r}_m,z_{jm},\beta(\succcurlyeq))| < \gamma$ uniformly. Further, in the event $C = \{\boldsymbol{\varepsilon} \mid |\varepsilon_j| \le \eta \text{ for } j \in \mathbf{J}\}$ that has $\text{Prob}(C) > 1 - \gamma/2$, $|\widetilde{V}(I,\mathbf{r}_m,z_{jm},\succcurlyeq) - V(I,\mathbf{r}_m,z_{jm},\beta(\succcurlyeq),\sigma(\succcurlyeq),\varepsilon_j)| < \gamma$ uniformly.

(ii) Suppose $\tilde{\delta}_{km}(I,\succcurlyeq) \equiv \tilde{\delta}_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq)$ is the choice indicator given by (14) for $\widetilde{V}$, and let $\delta_{km}(I,\beta,\sigma) \equiv \delta_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\beta,\sigma,\boldsymbol{\varepsilon})$ be an indicator for the discrete alternative that maximizes $V(I - p_{jm},\mathbf{r}_m,z_{jm},\beta,\sigma,\varepsilon_j)$ on $\mathbf{J}_m$. Then except for $\succcurlyeq$ and $\boldsymbol{\varepsilon}$ each in sets that have probability at most $\gamma/3$, $\delta_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\beta(\succcurlyeq),\sigma(\succcurlyeq),\boldsymbol{\varepsilon}) = \tilde{\delta}_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq)$. Letting $\widetilde{P}_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,s)$ denote the true discrete choice probability, from (16), and $P_k(I_m,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,s) = \mathbf{E}\delta_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\beta,\sigma,\boldsymbol{\varepsilon}) \equiv \mathbf{E}_{\beta,\sigma|s} \frac{\exp(v_{km}(I,\beta)/\sigma)}{\sum_{j\in J_m}\exp(v_{jm}(I,\beta)/\sigma)} \equiv \int_{\mathcal{B}\times\mathcal{S}} \frac{\exp(v_{km}(I,\beta)/\sigma)}{\sum_{j\in J_m}\exp(v_{jm}(I,\beta)/\sigma)} F(d\beta,d\sigma|s)$, where $v_{km}(I,\beta) \equiv X(I,\mathbf{r}_m,z_{km}) \cdot \beta - p_{km}$, one then has, uniformly, $|\widetilde{P}_k(I,\mathbf{p}_m,\mathbf{r},\mathbf{z}_m,s) - P_k(I,\mathbf{p}_m,\mathbf{r},\mathbf{z}_m,s)| < \gamma$.

(iii) Let $F(A|s) \equiv F^H(\{\succcurlyeq \in H \mid (\beta(\succcurlyeq),\sigma(\succcurlyeq)) \in A\}|s)$ for Borel sets $A \subseteq \mathcal{B}\times\mathcal{S}$ and $s \in S$, and let $F_T(A|s)$ denote the empirical probability obtained from T independent draws from F. Let $\mathcal{F}_1$ denote the family of functions of the form (17) for $j \in \mathbf{J}_m$, $\mathcal{F}_2$ denote the family of functions formed as differences of the functions in $\mathcal{F}_1$, and $\mathcal{F}$ denote the family of functions of the form $\min(f_1,\ldots,f_K)$ for $f_k \in \mathcal{F}_2$ and $1 \le K \le |\mathbf{J}|$, plus the function $f \equiv 1$. Let $\mathcal{K}$ denote the family of functions $\exp(v(I,p_{jm},\mathbf{r}_m,z_{jm},\beta)/\sigma)/\sum_{i\in J_m}\exp(v(I,p_{im},\mathbf{r}_m,z_{im},\beta)/\sigma)$ for v given in (17). Let $\mathcal{I}$ denote the family of indicator functions $i = \mathbf{1}(f>0)$ for $f \in \mathcal{F}$, and $\mathcal{G}$ denote the family of functions of the form $i \cdot f$ for $i \in \mathcal{J}$ and $f \in \mathcal{F}$. Letting $\mathbf{E}_{\beta,\sigma}$ and $\mathbf{E}_{\beta,\sigma,T}$ denote expectation operators with respect to F and F_T respectively, there exists T such that $\text{Prob}(\sup_{T' \ge T} \sup_{f\in\mathcal{F}\cup\mathcal{K}\cup\mathcal{I}\cup\mathcal{G}} |(\mathbf{E}_{T'} - \mathbf{E})f| > \gamma/3) < \gamma/3$.

(iv) Let $\widetilde{D}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq)$ and $D(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\beta,\sigma,\boldsymbol{\varepsilon})$ denote the continuous good demands given by (15) for the indirect utility functions $\widetilde{V}$ and V respectively. If on a closed subset A of $[I^L - p^U, I^U] \times R \times Z$, V is continuously differentiable in (I,r), then X can be selected with a sufficient number of terms so that on the set A and except for sets of $\succcurlyeq$ and $\boldsymbol{\varepsilon}$ that each have probability at most $\gamma/3$, $|\widetilde{D}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\succcurlyeq) - D(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\beta(\succcurlyeq),\sigma(\succcurlyeq),\varepsilon_j)| < \gamma$ uniformly.

Proof: Let $H^{\#}_{\delta}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m) = \cup_{k\in\mathbf{J}_m}\{\succcurlyeq \in H \mid \widetilde{V}(I - p_{km},\mathbf{r}_m,z_{km},\succcurlyeq) > \widetilde{V}(I - p_{jm},\mathbf{r}_m,z_{jm},\succcurlyeq) + \delta \text{ for } j \in \mathbf{J}_m \;\&\; j \ne k\}$ for $0 < \delta \le \gamma$. Then $H^{\#}_{\delta}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m) \searrow H^{\#}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)$, and A4 implies that there exists $\delta(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m) > 0$ such that $F(H^{\#}_{\delta(I,\mathbf{p}_m,\mathbf{r},\mathbf{z}_m)}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)|s) \ge 1 - \gamma/2$. Further, the continuity of $\widetilde{V}$ on $[I^L - p^U, I^U] \times R \times Z \times H$ implies there exists an open neighborhood $N(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)$ in $[I^L, I^U] \times P^{|\mathbf{J}_m|} \times R \times Z^{|\mathbf{J}_m|}$ such that $\succcurlyeq \in H^{\#}_{\delta(I,\mathbf{p}_m,\mathbf{r},\mathbf{z}_m)}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)$ and $(\tilde{I},\tilde{\boldsymbol{p}}_m,\tilde{\boldsymbol{r}},\tilde{\boldsymbol{z}}_m) \in N(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)$ imply $\max_{k\in\mathbf{J}_m}\{\widetilde{V}(\tilde{I} - \tilde{p}_{km},\tilde{\mathbf{r}},\tilde{z}_{km},\succcurlyeq) - \max_{j\ne k}\widetilde{V}(\tilde{I} - \tilde{p}_{jm},\tilde{\mathbf{r}},\tilde{z}_{jm},\succcurlyeq)\} > \delta(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)/2$. One can then extract a finite family of these open neighborhoods that cover $[I^L, I^U] \times P^{|\mathbf{J}_m|} \times R \times Z^{|\mathbf{J}_m|}$. Let $\delta_0 > 0$ denote the minimum of the $\delta(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)$ for this finite family and define a <u>constant</u> $\sigma = \sigma(\succcurlyeq) \equiv \frac{\delta_0}{12\eta}$. Recall that Z is a finite union of disjoint rectangles. Combine each of these rectangles with the rectangular domains of income and prices, shift and scale these rectangles so they form a unit cube, and apply Appendix Theorem A.1 to establish the existence of a vector of multivariate polynomials $v(I,\mathbf{r}_m,z_{jm},\succcurlyeq) \equiv X(I,\mathbf{r}_m,z_{jm}) \cdot \beta(\succcurlyeq)$ that satisfy $|\tilde{v}(I,\mathbf{r}_m,z_{jm},\succcurlyeq) -$

$v(I,\mathbf{r}_m,z_{jm},\succcurlyeq)| < \frac{\delta_0}{12} \leq \gamma$. From the properties of EV1 variates, the event C has Prob(C) > 1 – γ/2, and if C, then $|\sigma\varepsilon_j|$ $< \delta_0/12$. In the event C, $V(I,\mathbf{r}_m,z_{jm},\succcurlyeq)$ given by (17), (i) is established by

$$|\widetilde{V}(I,\mathbf{r}_m,z_{jm},\succcurlyeq) - V(I,\mathbf{r}_m,z_{jm},\succcurlyeq)| \leq \left|v(I,\mathbf{r}_m,z_{jm},\succcurlyeq) - \tilde{v}(I,\mathbf{r}_m,z_{jm},\succcurlyeq)\right| + |\sigma\varepsilon_j| < \delta_0/6.$$

For any point $(\tilde{I},\tilde{\boldsymbol{p}}_m,\tilde{\boldsymbol{r}},\tilde{\boldsymbol{z}}_m) \in [I^L,I^U] \times P^{J_m+1} \times R \times Z^{J_m+1}$, let $(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)$ be the center of a neighborhood in the open cover that includes $(\tilde{I},\tilde{\boldsymbol{p}}_m,\tilde{\boldsymbol{r}},\tilde{\boldsymbol{z}}_m)$. The probability of the event $C \cap H^{\#}_{\alpha(I,\mathbf{p}_m,\mathbf{r},\mathbf{z}_m)}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)$ is at least 1 – γ. In this case, $V(\tilde{I} - \tilde{p}_{km},\tilde{\mathbf{r}},\tilde{z}_{km},\succcurlyeq) - \max_{j\neq k} V(\tilde{I} - \tilde{p}_{jm},\tilde{\mathbf{r}},\tilde{z}_{jm},\succcurlyeq) > \frac{\delta(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m)}{2} > \delta_0/2$ for some k implies $\widetilde{V}(\tilde{I} - \tilde{p}_{km},\tilde{\mathbf{r}},\tilde{z}_{km},\succcurlyeq) - \max_{j\neq k} \widetilde{\mathbf{V}}(\tilde{I} - \tilde{p}_{jm},\tilde{\mathbf{r}},\tilde{z}_{jm},\succcurlyeq) > \delta_0/6$. Then (ii) holds with probability at least 1 – γ. The bound on the difference between the exact and approximate choice probabilities then follows.

The proof of (iii) utilizes results on convergence of empirical processes given in Appendix A. The functions in the family $\mathcal{F}_1$, and hence in $\mathcal{F}_2$, are linear in $(\beta,\gamma,\varepsilon_0,\ldots,\varepsilon_{|J|})$. Then these families are contained in a finite-dimensional subspace defined by their intercepts and slope coefficients. The functions in $\mathcal{F}_2$ are Lipschitz in these intercepts and slope coefficients, implying that $\mathcal{F}$ is Lipschitz in these parameters. By construction, the domain $[I^L-p^U,I^U]\times R\times Z\times\mathcal{B}\times\mathcal{S}$ and the domain $\mathcal{B}\times[\sigma^L,\sigma^U]$ of (β,σ) are compact, so that $v(I,p_{jm},\mathbf{r}_m,z_{jm},\beta)$ is bounded on its domain by a constant M. Therefore, $f^* = M + \sigma^U|\varepsilon|$ is an envelope function for $\mathcal{F}_1$ and $2f^*$ is an envelope function for $\mathcal{F}_2$, and hence for $\mathcal{F}$, that from Appendix B(a) satisfies $\mathbf{E}_F f^* \leq M + 1.219384\,\sigma^U$. The family $\mathcal{K}$ is Lipschitz in $(\beta,\sigma) \in \mathcal{B}\times[\sigma^L,\sigma^U]$ since $\sigma^L > 0$, with envelope function $f^* \equiv 1$. Apply Theorem A.3 to establish the result for $\mathcal{F}$ and $\mathcal{K}$, and Theorem A.4 to establish the result for $\mathcal{I}$ and $\mathcal{G}$.

From Theorem 2.5, $\widetilde{V}$ is almost everywhere continuously differentiable in $(I,\mathbf{r})$, and where it is, continuous good demands are unique and are given by (15). Let A be a closed set on which this continuous differentiability holds. Then, Lemma 6 establishes that the derivatives of V approximate uniformly on A the corresponding derivatives of $\widetilde{V}$. Combined with the bi-Lipschitz property of $\widetilde{V}$ in I, this establishes (v). ■

The additive EV1-distributed disturbances ε_j in (17) are introduced as a mathematical convenience, perturbations that for small positive $\sigma(\succcurlyeq)$ smooth expected utility and give choice probabilities that are well-behaved mixed multinomial logits, while approximating closely the choices and continuous good demands from the underlying true utility. Under Assumptions A1-A4, the approximation properties (i), (iii), and (iv) of Theorem 2.7 continue to hold even in the limit $\sigma(\succcurlyeq) \equiv 0$. The EV1 and independence assumptions on the ε_j are not essential for smoothing the choice probabilities; any absolutely continuous distribution with well-behaved tails accomplishes this. The ε_j in (17) can be treated as predetermined at the time of consumer choice, and thus independent of income transfers or market scenario, so that (17) is consistent with fully neoclassical consumer behavior. In the literature on discrete choice, the ε_j are often characterized as random perturbations or tremble in individual utility, and attributed to the limits of psychophysical discrimination, as in Thurstone (1927). This relaxation of neoclassical assumptions can be made more general by allowing the individual consumer's preference preorder $\succcurlyeq$ to be a random draw from H in each choice situation, perhaps representable as tremble centered on a core preference preorder. In this case, F^H is a convolution of population heterogeneity in core preferences and individual preference tremble. The implications of true preference tremble for demand analysis and the welfare calculus are deferred to the section on decision versus experienced utility.

Result (iii) in this theorem shows that utility and the distribution of tastes can be approximated using the empirical distribution of the finite-dimensional taste parameters (β,σ). This can be interpreted as a member of the finite-parameter family that places probability at each support point. Obviously, it is then possible to achieve an approximation of the same precision using other finite-parameter families of distributions with the same number of parameters. The proof of this theorem assumes a constant for the scaling factor σ, and this is sufficient for the approximation results, but allowing heterogeneity in σ may allow more parsimonious approximations with respect to the specification of β.

The direct utility function (5) given by Lemma 2 can be interpreted as a continuous mapping from the compact space of preferences H onto a compact subset $\mathcal{U}$ of the normed linear space $\mathfrak{C}(Q'\times Z)$ of continuous real-valued functions $u:Q'\times Z \longrightarrow \mathbb{R}$. The probability F^H on H induces a probability F^U on $\mathcal{U}$ that satisfies $F^U(A|s) = F^H(\{\succcurlyeq\in H|U(\cdot,\cdot,\succcurlyeq)\in A\} \mid s)$ for measurable subsets A of $\mathcal{U}$. Then, given A1-A4, the field of preferences can be characterized by $(\mathcal{U},F^U)$ rather than (H,F^H). With this characterization, the money-metric utility function (9) in Lemma 4 is written $V:[I^L-p^U,I^U]\times R\times Z\times\mathcal{U} \longrightarrow \mathbb{R}$, and correspondingly the choice indicator $\delta_k(I,\mathbf{p}_m,\mathbf{r_m},\mathbf{z}_m,U)$ from (14) and continuous good demand $D(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,U)$ from (15) are written as functions of $U \in \mathcal{U}$.

Consider the family (17) of the indirect utility functions $V(I_m - p_{jm},\mathbf{r}_m,z_{jm},\beta,\sigma,\varepsilon_j) \equiv I_m + v(I_m,p_{jm},\mathbf{r}_m,z_{jm},\beta) + \sigma\varepsilon_j$, where $v(I_m,p_{jm},\mathbf{r}_m,z_{jm},\beta) \equiv X(I_m - p_{jm},\mathbf{r}_m,z_{jm})\cdot\beta - p_{jm}$. A major simplification of the welfare calculus occurs when $v(I,p_{jm},\mathbf{r}_m,z_{jm},\beta)$ is independent of income. Make explicit the price index $\pi = \pi(\mathbf{r}_m)$ used to deflate income and prices to real terms, where $\pi(\mathbf{r}_a) = \pi(\mathbf{r}_b)$ by assumption, and rewrite V as

$$(18) \qquad V(I_m - p_{jm},\mathbf{r}_m,z_{jm},\beta,\sigma,\varepsilon_j) = I_m/\pi(\mathbf{r}_m) + v(p_{jm}/\pi(\mathbf{r}_m),\mathbf{r}_m/\pi(\mathbf{r}_m),z_{jm},\beta) + \sigma\varepsilon_j.$$

But this is a Gorman Polar Form, with the properties that choice among the products $j \in \mathbf{J}_m$ is independent of income, and continuous good demands have the form

(19) $\qquad D(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\beta,\sigma,\boldsymbol{\varepsilon}_m) = -\partial v(p_{jm}/\pi(\mathbf{r}_m),\mathbf{r}_m/\pi(\mathbf{r}_m),z_{jm},\beta)/\partial\mathbf{r}_m + [I_m + v(p_{jm}/\pi(\mathbf{r}_m),\mathbf{r}_m/\pi(\mathbf{r}_m),z_{jm},\beta)]\partial\ln\pi(\mathbf{r}_m)/\partial\mathbf{r}_m$,

so that the only goods showing income effects are those whose prices influence the index π, and the Engle curves for these goods are affine linear. The Gorman polar preference field has been studied extensively in welfare economics, and has important aggregation properties for both continuous and discrete choice; see Chipman and Moore (1980,1990), Small and Rosen (1981), and McFadden (2004,2014). The Gorman form (19) defines a *hedonic preference field* in which product attributes influence tastes only through an *effective price,* $\tilde{p}_{jm} = p_{jm} - X(\mathbf{r}_m,z_{jm})\cdot\beta$.

The approximation (17) is consistent with the approach to welfare analysis taken by Jorgenson (1997) using translog utility function families with parameterized observed heterogeneity. Other empirical demand analysis systems, such as generalized Gorman (Blackorby et al, 1978) or Deaton-Muellbauer (1980), can also be interpreted as specializations of (17). The money-metric property imposed on (17) will in general put side restrictions on the parameters of these functional families. These are most easily handled in applications by specifying (17) without the money-metric restriction, and then obtaining the marginal utility of income from these forms that can later be used to convert utility differences to (approximate) money-metric terms.

3. WELFARE ANALYSIS

We restate for product markets the neoclassical welfare calculus outlined in Section 1, utilizing the treatment of consumer theory given in Section 2.[8] There is a baseline, "as is," or "default" policy/scenario m = *a* and a counterfactual, "but for," or "replacement" policy/scenario m = *b*.[9] Consumers face menus of mutually exclusive products $j \in \mathbf{J}_m \subseteq \mathbf{J}$ with at least one "benchmark" or "no purchase" alternative whose attributes are unaffected by scenario changes.[10] Our analysis will be carried out for a population of consumers who are neoclassical maximizers of preferences that satisfy Assumptions A1-A4 and that are predetermined and unaffected by income

[8] The basics of this theory can be found in Varian (1992, Chap. 7, 10), Mas-Colell, Whinston, and Green (1995, Chap. 3), and other graduate-level textbooks. See also McFadden and Winter (1966) and Border (2014).

[9] For convenience we will use the "baseline/as is/default" and "counterfactual/but for/replacement" labels for both retrospective analysis of past policy and prospective analysis of policies not yet implemented, noting that these labels are arbitrary and interchangeable in many prospective applications. In retrospective applications, associating *a* with the historical scenario and *b* with the counterfactual leads to measures of welfare change often termed "Willingness to Pay" (WTP), while reversing these labels and making *b* the baseline leads to "Willingness to Accept" (WTA) welfare measures.

[10] If the products in an application are not mutually exclusive, or the consumer can buy more than one unit of a product, then **J** indexes the mutually exclusive possible portfolios of product purchases. In general, **J** may index locations or "addresses" in physical or hedonic space, and with added technical machinery is not restricted to be finite.

transfers or scenario changes that alter market opportunities.[11] These consumers have indirect decision-utility functions that from Theorem 2.7 are uniformly approximated by

$$(20) \qquad V(I - p_{jm},\mathbf{r}_m,z_{jm},\beta,\sigma,\varepsilon_j) \equiv I + v_{jm}(I,\beta) + \sigma\varepsilon_j,$$

where $v_{jm}(I,\beta)$ is shorthand for $v(I,p_{jm},\mathbf{r}_m,z_{jm},\beta) \equiv X(I - p_{jm},\mathbf{r}_m,z_{jm})\beta - p_{jm}$. The vector β and positive scalar σ are randomly distributed in the population with a probability $F(\beta,\sigma|s,\alpha)$ that is in a parametric family with parameter α, given observed socioeconomic history s, and the ε_j are independent standard Extreme Value type I random variables. As discussed earlier, the ε_j are introduced as a mathematical convenience, but will often be interpreted as contributions of unobserved perceptions and attributes to the utility of product j. By construction, V is money-metric for a "no purchase" or "benchmark" alternative in scenario a (e.g., $v(I,p_{0a},\mathbf{r}_a,z_{0a},\beta) \equiv 0$). In the event H_m^k = $\{\boldsymbol{\varepsilon} \mid V(I - p_{km},\mathbf{r}_m,z_{km},\beta,\sigma,\varepsilon_k) > V(I - p_{jm},\mathbf{r}_m,z_{jm},\beta,\sigma,\varepsilon_j) \text{ for } j \in \mathbf{J}_m\backslash\{k\}\}$, the consumer maximizes (20) at alternative $k \in \mathbf{J}_m$, an event indicated by $\delta_{km}(I,\beta,\sigma,\boldsymbol{\varepsilon}) = 1$, with a probability that is uniformly approximated by a mixed multinomial logit (MMNL),

$$(21) \qquad P_{km} = P_k(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,,s,\alpha) \equiv \mathbf{E}_{\beta,\sigma|s}\, L_{km}(I,\beta,\sigma)\ ,$$

where $L_{km}(I,\beta,\sigma) = \frac{\exp(v_{km}(I,\beta)/\sigma)}{\sum_{j\in\mathbf{J}_m}\exp(v_{jm}(I,\beta)/\sigma)}$ is the "flat" multinomial logit probability of $\delta_{km}(I,\beta,\sigma,\boldsymbol{\varepsilon}) = 1$ given (β,σ). The components of the vector of taste parameters β are termed "part-worth" or Willingness-to-Pay (WTP) coefficients for unit changes in the corresponding components of X. Unconditional maximum indirect utility in scenario m when income is I , our policy-independent yardstick for welfare analysis, then satisfies

$$(22) \qquad u_m = \mathcal{V}(I,\mathbf{p}_m,\mathbf{r}_m,\mathbf{z}_m,\beta,\sigma,\boldsymbol{\varepsilon}) \ \equiv \max_{j\in\mathbf{J}_m}\ [I + v(I,p_{jm},\mathbf{r}_m,z_{jm},\beta) + \sigma\varepsilon_j].$$

There are three substantive questions whose resolution affects the form of the welfare calculus: (1) Is the analysis *prospective*, comparing policies not yet put into place, or *retrospective*, comparing "as is" and "but for" past policies? (2) Is information on the tastes of individual consumers *complete* or *partial*, and if partial what welfare measures are relevant to transfers that can actually be fulfilled? (3) Should well-being be assessed in terms of the *decision-utility* postulated to determine economic demand behavior, or in terms of *experienced-utility* after taste ambiguities and uncertainties are resolved? These questions are discussed in Sections 4, 5, and

[11] Our analysis lumps unobserved perceptions and attributes of alternatives together with unobserved preferences. To maintain taste invariance when these unobserved factors are influenced by policy, we would need to make these sources of randomness explicit and consider how to detect their presence and identify their influence on welfare.

6 below. In the remainder of this section, we restate for our general model of discrete product choice and neoclassical assumptions the welfare measures introduced in Section 1.

A standard welfare measure for the net gain in well-being from scenario *b* relative to scenario *a* is *Hicksian Compensating Variation* (HCV), the net decrease in scenario *b* income that makes the two scenarios indifferent. Let HCV(s,k,β,σ,$\boldsymbol{\varepsilon}$) denote this measure for an observed history *s* and scenario *a* choice k, and a vector (β,σ,$\boldsymbol{\varepsilon}$) of unobservables.[12] In terms of the conditional indirect utility function (21), HCV(s,k,β,σ,$\boldsymbol{\varepsilon}$) satisfies

$$(23)\quad \max_{j\in\mathbf{J}_b} \left[I_b - \mathrm{HCV} + \mathrm{v}(I_b - \mathrm{HCV}, p_{jb}, \mathbf{r}_b, z_{jb}, \beta) + \sigma\varepsilon_j\right] = I_a + \mathrm{v}(I_a, p_{ka}, \mathbf{r}_a, z_{ka}, \beta) + \sigma\varepsilon_k.$$

Removing the conditioning on k when it is not observed gives the measure HCV(s,β,σ,$\boldsymbol{\varepsilon}$) = $\min_{k\in\mathbf{J}_a} \mathrm{HCV}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})$. Another standard welfare measure, *Hicksian Equivalent Variation* (HEV), denoted HEV(s,β,σ,$\boldsymbol{\varepsilon}$), is the net increase in scenario *a* income that makes the two scenarios indifferent. Because the scenario *a* choice may change with the income transfer, HEV does not depend explicitly on the uncompensated scenario a choice k; from (23), HEV(s,β,σ,$\boldsymbol{\varepsilon}$) satisfies

$$(24)\quad \max_{j\in\mathbf{J}_b} \left[I_b + \mathrm{v}(I_b, p_{jb}, \mathbf{r}_b, z_{jb}, \beta) + \sigma\varepsilon_j\right] = \max_{j\in\mathbf{J}_a} \left[I_a + \mathrm{HEV} + \mathrm{v}(I_a + \mathrm{HEV}, p_{ja}, \mathbf{r}_a, z_{ja}, \beta) + \sigma\varepsilon_j\right].$$

Sometimes, HCV is termed *Willingness-to-Pay* (WTP), and HEV is termed *Willingness-to-Accept* (WTA); this terminology is related to the description of β as a vector of WTP coefficients, but only in special cases is there a simple mapping between β and HCV or HEV.

The definition of *Market Compensating Equivalent* (MCE) that generalizes (2) is the difference in the utilities (23) that the consumer would attain in scenarios *a* and *b* in the absence of compensation, scaled by a MUI in scenario *a* scaling the utility difference in monetary units. The conditional indirect utility $\mathrm{V}(I - p_{jm},\mathbf{r}_m,z_{jm},\beta,\sigma,\varepsilon_j) \equiv I + \mathrm{v}(I,p_{km},\mathbf{r}_m,z_{km},\beta) + \sigma\varepsilon_{km}$ is denominated in monetary units and is money-metric for a "no purchase" alternative. However, if there are neoclassical income effects for other alternatives k, one has $\partial\mathrm{V}(I - p_{km},\mathbf{r}_m,z_{km},\beta,\sigma,\varepsilon_k)/\partial I \equiv 1 + \partial\mathrm{v}(I,p_{km},\mathbf{r}_m,z_{km},\beta)/\partial I \not\equiv 1$. If alternative k is chosen in scenario *a*, then

$$(25)\quad \mathrm{MCE}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon}) = \frac{\max_{j\in\mathbf{J}_b}\left[I_b + \mathrm{v}(I_b,p_{jb},\mathbf{r}_b,z_{jb},\beta) + \sigma\varepsilon_j\right] - \max_{j\in\mathbf{J}_a}\left[I_a + \mathrm{v}(I_a,p_{ja},\mathbf{r}_a,z_{ja},\beta) + \sigma\varepsilon_j\right]}{\mu_k(I_a,\beta)}.$$

Where the MUI,

---

[12] Note that when k is observed in scenario *a*, the distribution of $\boldsymbol{\varepsilon}$ is conditioned on the event $\{\boldsymbol{\varepsilon}\,|\,\delta_{ka}(I,\beta,\sigma,\boldsymbol{\varepsilon}) = 1\}$.

(26) $\mu_k(I_a,\beta) = \partial v(I_a - p_{ka},\mathbf{r}_a,z_{ka},\beta)/\partial I_a \equiv 1 + [\partial X(I_a - p_{ka},\mathbf{r}_a,z_{ka})/\partial I_a]\beta$,

gives a definition of MCE(s,k,β,σ,$\boldsymbol{\varepsilon}$) that at least locally has the money-metric property in scenario a.[13] Later, when we consider cases where choice k in scenario a is not observed, or one observes or uses only the information that "as is" choice is from a set $\mathbf{D} \subseteq \mathbf{J}_m$, $\mu_k(I_a,\beta)$ will be replaced by a scale factor $\mu_D(I_a,\beta)$. Note that $\mu_k(I_a,\beta)$ defined by (26) is independent of σ and $\boldsymbol{\varepsilon}$, and if X does not depend on income, then $\mu_k(I_a,\beta) \equiv 1$.

The measure (25) can be interpreted as a generalization to multiple products with varying attributes of the Marshallian consumer surplus (MCS) introduced in Section 1; this generalization is more convenient for applications than multivariate extensions, path-dependent when there are income effects, of the integral form (2) for MCS. First-order Taylor's expansions of (24) imply

$$(27)\quad \mathrm{HCV}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})\cdot\mu' = \mathrm{HEV}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})\cdot\mu'' = \mathrm{MCE}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})\cdot\mu_{\mathrm{k}}(I_{\mathrm{a}},\beta)\ ,$$

where $\mu' = 1 + \partial v(I' - p_{jb},\mathbf{r}_b,z_{jb},\beta)/\partial I$ and $\mu'' = 1 + \partial v(I'' - p_{ka},\mathbf{r}_a,z_{kb},\beta)/\partial I$ are MUI at the chosen alternatives j in scenario b and k in scenario a, respectively, when there is no compensation, evaluated at incomes I' and I'' intermediate between uncompensated and compensated levels. The measures HCV, HEV, and MCE all agree on sign, but in general can differ in magnitude. However, if the marginal utility of income is constant, then HCV = HEV = MCE. In general, (24) can be solved quickly for HCV or HEV by iteration starting from MCE.

It is common in applied welfare analysis to aggregate money-metric measures of individual benefits from a policy change, net of allocated costs and fulfilled transfers, and judge the change desirable if this aggregate welfare measure exceeds unallocated costs. Ideally, the cost allocation and fulfilled transfers exhaust the feasible opportunities for socially desirable income redistribution, so that the feasibility-constrained social marginal utilities of income for consumers are the same and equal weighting of consumers in the aggregate welfare criterion is appropriate. Restrictions on the nature of the preference field, the set of policies under consideration, and/or the measure of individual welfare are required for the aggregate welfare criterion to order policies and identify a best policy; otherwise, it may fail to satisfy the irreflexivity or transitivity conditions required of an order.

[13] The scale factor $\mu_k(I_a,\beta)$ in the definition of MCE is natural for retrospective analysis where the consumer has experienced scenario a, or for prospective analysis when scenario a is a default that will occur unless there is a policy intervention, and is unambiguous when indirect utility has been transformed so that the marginal utility of income in scenario a remains constant when income changes. However, more generally, MCE will be affected by transformations of utility and the evaluation point for the marginal utility of income, and additional criteria may be needed to select among alternative versions of MCE.

When the aggregate criterion does order policies, it has the properties of a Bergson (1938) social welfare function, and is then subject to the challenges of social choice theory; see Arrow (1950), Harsanyi (1955), Sen (2017, p. 385).

If the set of policies under consideration along with their accompanying fulfilled transfers are Pareto-ordered, then the aggregate welfare criterion with any of the transfers HCV, HEV, or MCE will follow the same order. Alternately, when marginal utilities of income are constant over the domain of consumption induced by the policy set, and equal across individuals, the aggregate welfare criterion with MCE and any pattern of transfers orders policies; this case corresponds to a preference field of Gorman polar form with parallel income-expansion paths; see Chipman and Moore (1980,1990), McFadden (2004). Kaldor (1936) and Scitovsky (1942) give an argument that suggests the aggregate welfare criterion using HCV, termed the Kaldor-Hicks criterion, orders policies, and that this is a basis for preferring HCV over MCS. On closer examination, this argument holds only in cases such as constant, equal marginal utilities of income, or policies incorporating transfers that are Pareto-ordered, in which case, either HCV or MCE can be used. Otherwise, the aggregate welfare criterion with either HCV or MCE may fail to order the policy alternatives.

With the apparatus above, practical welfare analysis of product markets can be carried out in three steps. First, observations on the market choices of surveyed consumers, augmented by extra-market data on stated preferences if necessary to identify tastes for relevant attributes, can be used to estimate the mixed MNL model (22) and recover the probability $F(\beta,\sigma|s,\alpha)$. An obvious caution is that the vector of predetermined functions X in (21) has to be comprehensive enough to achieve the approximation accuracy promised by Theorem 2.7, so that estimation of (22) needs to include a careful econometric specification analysis. A "method of sieves" approach to the specification of X provides practical guidelines for this specification search. With this caveat, this setup is both practical and sufficiently general to handle welfare analysis of policy changes that affect discrete choice without making unwarranted assumptions on preferences.

Second, construct a large synthetic population. Start from a random sample from the target population. For each sampled person, assign a history s, incomes I_a and I_b, choice sets $\mathbf{J}_a$ and $\mathbf{J}_b$, and market environments $(\mathbf{p}_a,\mathbf{r}_a,\mathbf{z}_a)$ and $(\mathbf{p}_b,\mathbf{r}_b,\mathbf{z}_b)$, using available data for the sampled individual wherever possible in order to preserve ecological correlations in the target population. Make multiple draws of (β,σ) from the estimated probability $F(\beta,\sigma|s,\alpha)$ and of $\boldsymbol{\varepsilon}$ from the standard Extreme Value Type I distribution. Assign utility-maximizing choices k in scenario *a* and j in scenario *b*. Each draw defines a synthetic consumer.

Third, calculate the measures $\mathrm{HCV}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})$, $\mathrm{HEV}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})$, and $\mathrm{MCE}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})$ from (24) and (25) for each consumer in the synthetic population. These measures can be aggregated over this synthetic population or subpopulations to estimate hypothetical compensating transfers for relevant consumer classes. However, transfers that are actually fulfilled in the target population can depend only on observable history *s* and (if observed) the scenario *a* choice k. Define uniform transfers UMCE(*s*,k) = $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,k}\mathrm{MCE}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})$ and UMCE(*s*) = $\mathbf{E}_{k,\beta,\sigma,\boldsymbol{\varepsilon}|s}\mathrm{MCE}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})$. Fulfillment of these transfers in the real population in retrospective welfare analysis will not in general make individual consumers "whole", but will balance individual gains and losses in the sense that a MCE welfare measure taken <u>subsequent</u> to these uniform transfers aggregates to zero. In the same way, one can solve for uniform transfers t_k = UHCV(*s*,k) and t = UHCV(*s*) that if fulfilled in scenario *b* balance the gains and losses from the remaining unfulfilled Hicksian transfers, so that a subsequent MCE aggregates to zero:

$$(28)\qquad 0 = \begin{cases} \mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,k}\left\{\max\limits_{j\in\mathbf{J}_b}\ \left[I_b - t_k + v\big(I_b - t_k, p_{jb}, \mathbf{r}_b, z_{jb}, \beta\big) + \sigma\varepsilon_j\right] - \left[I_a + v(I_a, p_{ka}, \mathbf{r}_a, z_{ka}, \beta) + \sigma\varepsilon_k\right]\right\} \\ \mathbf{E}_{k,\beta,\sigma,\boldsymbol{\varepsilon}|s}\left\{\max\limits_{j\in\mathbf{J}_b}\ \left[I_b - t + v\big(I_b - t, p_{jb}, \mathbf{r}_b, z_{jb}, \beta\big) + \sigma\varepsilon_j\right] - \left[I_a + v(I_a, p_{ka}, \mathbf{r}_a, z_{ka}, \beta) + \sigma\varepsilon_k\right]\right\} \end{cases}.$$

Analogous definitions can be given for UHEV(*s*,k) and UHEV(*s*).[14] Note that the measures considered in this paragraph are all based on predetermined decision-utility, with no adjustment for possible tremble in decision utility or differences in decision and experienced utility.

4. PROSPECTIVE VERSUS RETROSPECTIVE WELFARE ANALYSIS

Traditional welfare theory considers a prospective policy change in a static "what if" environment. An "incumbent" or "default" policy/scenario *a* is compared with a "replacement" policy/scenario *b* in a situation where neither has been implemented and both are on the table. The theory assumes that the policymaker has the information and authority to carry out net lump sum transfers in the event that policy *b* is adopted, adjusted for direct policy-induced effects on incomes, that make each consumer indifferent between the policies, and assumes that if policy *b* is adopted, these transfers are fulfilled <u>before</u> consumers maximize utility. Under these conditions, the Hicksian Contingent Variation (HCV) defined in (24) is the precise measure of each lump sum transfer required. If instead, *a* and *b* are reversed, so that transfers are fulfilled if *a* is adopted, then the Hicksian Equivalent Variation (HEV) is the precise measure of each lump sum transfer required. So long as population

[14] Another approach to defining UHCV(*s*,k) and UHEV(*s*,k) is to consider "representative" utility for the class of consumers with history *s*, perhaps the expectation of (21) with respect to the unobservables, and then define UHCV(*s*,k), or UHEV(*s*,k) as analogs of HCV or HEV for "representative" utility. However, these definitions will not in general have the property that UHCV(*s*,k) = $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,k}$HCV(β,σ,**ε**) or UHEV(*s*,k) = $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,k}\mathrm{HEV}(\beta,\sigma,\boldsymbol{\varepsilon})$.

aggregate HCV or HEV, adjusted for policy-induced income changes, exceeds zero, a shift from policy *a* to policy *b* with the exact individual transfers fulfilled, plus any distribution of the residual surplus, is a Pareto improvement.

In practice many welfare calculations are *retrospective* rather than prospective. The welfare question is what transfers *after the fact* redress harm from a past "as is" or "baseline" scenario *a* in which some products were defective or improperly marketed, using as a benchmark a "but for" or "counterfactual" scenario *b* in which these flaws would have been absent.[15] A key feature of these applications is that the transfer occurs after the decision-utility-maximizing choice would have been made in the "but for" scenario, and hence these transfers could not be a factor in "but for" choice. Put another way, the "but for" utility maximization that would have occurred at the consumer's original income will not in general coincide with that assumed in the Hicksian compensating variation calculation in which the transfer would have been made prior to consumer choice and would have influenced that choice. Since at the time the corrective transfer is being considered, the consumer is in the "as is" situation, this transfer is denominated in "as is" monetary units. Then, the transfer that "makes whole the consumer with choice k in the baseline scenario" equals the difference in the utilities (21) that would have been attained in the "but for" and "as is" scenarios, scaled to "as is" monetary units, the MCE (25).

Suppose the purpose of a prospective policy analysis is not to actually fulfill the HCV or HEV transfers associated with a move from scenario *a* to scenario *b*, but simply to determine whether it is possible *in principle* to compensate consumers so that the move from scenario *a* to scenario *b* would be a Pareto improvement. Then, arguably, aggregate MCE rather than aggregate HCV or HEV is the appropriate welfare criterion. Further, MCE is easier to compute and aggregate than HCV or HEV, since it is obtained as an explicit solution (25) from the indirect utility functions (21) of individual consumers, and the distribution of these solutions in the target population. Equation (27) shows that HCV, HEV, and MCE differ only due to differences in the marginal utility of income at different arguments. Later, we show in examples that these differences are often but not always modest. Then, the distinction between prospective and retrospective welfare measures often will be empirically unimportant, but occasionally will be of practical as well as theoretical significance.

The distinction we have made between prospective and retrospective welfare analysis does not require explicit consumer dynamics, but a MCE transfer to redress past harm obviously occurs at some time later than the period

[15] Retrospective policy analysis is often conducted in conjunction with litigation, and statues and legal rulings often control the definition of harm and the scope and magnitude of remedies. These legal standards are often rooted in economic arguments, but may nevertheless deviate from a purely economic analysis of harm and remedy. In this paper, we consider only the economic foundation of retrospective analysis, and do not take up legal considerations.

of the harm, introducing issues such as discounting and pre-judgement interest, but more fundamentally the longer-run impacts of injury on consumer assets and opportunities. We leave this as a topic for future research, but note that in a fully dynamic model, the impact of policy on state variables justifies scaling MCE in monetary units that make the consumer whole in terms of lifetime well-being. [16]

5. PARTIAL OBSERVABILITY AND WELFARE AGGREGATES

Traditional welfare analysis assumes that the individual utility functions required to calculate measures of well-being can be recovered fully (with money-metric scaling) from observations on this consumer's market choices. This is unrealistic, first because the analyst typically has observations on a consumer's choices in only a small number of market environments, often only one, and because markets are observed only over a limited range of conditions. For example, variations in historical product prices are limited by production costs and competition between products, and the dimensionality of possible product attributes is high, with only a limited range of bundles of attributes appearing in historically available products. However, different consumers generally face somewhat different observed market environments, and if one can maintain the consumer sovereignty assumption that consumer tastes are predetermined at the time of market choice, and assume plausibly that given *s* there is no ecological correlation of market environments and tastes, then observations across consumers can be used to estimate the distribution of tastes in the population. Further, in many applications it is reasonable to assume that consumers value products using hedonic effective prices that adjust market price for the attributes of the product; then the analysis can recover distributions of hedonic weights. This will often be sufficient to infer the distribution of consumer utilities for new or modified products even if their specific configurations of attributes are novel.

A more challenging recovery problem arises when markets are incomplete, due to transaction costs, asymmetric information that causes market failure through adverse selection and moral hazard, or failure to

[16] Technically, retrospective welfare analysis should be conducted with a multi-period consumer model, with redress in the second period from harm in the first period. If the consumer has intertemporarly separable utility, then the ideal MCE measure satisfies $V_{1b}(I_1) + V_2(I_2 - MCE) = V_{1a}(I_1) + V_2(I_2)$, where V_1 and V_2 are indirect utilities for the respective periods, and non-income arguments in indirect utility are suppressed. Applying the first mean value theorem for integrals, MCE = $[V_{1b}(I_2) - V_{1a}(I_1)]/\mu_2$, where μ_2 is a marginal utility of income in the second period. But the consumer will allocate income between periods to equate marginal utilities of income (without accounting for MCE), so that μ_2 will to a first approximation equal $\mu(I_a,\beta,\sigma)$. Consequently, the MCE defined in (25) approximates the two-period ideal. Further analysis of intertemporal utility to sharpen the definition of MCE is left to the reader.

establish ownership and control of the distribution of some goods and services. For example, consumers cannot insure against some kinds of events, cannot directly purchase environmental amenities such as clean air and unpolluted beaches, and lack market opportunities that show their tastes for "existence goods" such as protecting endangered species or reducing global warming. If there is sufficient market redundancy, or if there are active margins where unmarketed and marketed goods are complements or substitutes, then it may be possible to recover indirectly preferences for unmarketed goods. For example, consumer preferences for environmental amenities are reflected in their willingness to travel to unpolluted beaches or move to neighborhoods with cleaner air. However, when preferences for unmarketed goods and services leave no market trace, they obviously cannot be recovered from market data. Experimental methods for directly eliciting stated preferences for these goods in hypothetical markets are successful in some marketing contexts, but sensitivity to context and framing can make experimental data unreliable; see Ben Akiva et al (2016), McFadden (2017), Miller et al. (2011). For the remainder of this section, we assume that there is sufficient market information to recover distributions of preferences in the population, and study the construction of aggregate measures of welfare. These aggregates may be sufficient for policy decisions, or sufficient to determine transfers that are judged appropriate to remedy harm to a class of consumers even if the compensation is not exact for each individual.

When a welfare analysis seeks to fulfill the transfers HCV, HEV, or MCE that in retrospective or prospective applications leave a class of consumers indifferent to the policy change, an obvious limitation is that an actual transfer to a consumer can be a function only of observed characteristics. It is common in applied welfare analysis to estimate welfare effects by postulating a *representative consumer* whose demands are close to the per capita market demands of a consumer class, calculating the transfer that keeps "representative" utility constant, and assuming that this per capita transfer could in principle be redistributed to keep the utility of each consumer in the class constant. A necessary and sufficient condition for the existence of a representative consumer meeting these conditions exactly is that the utilities of individuals in the class be representable in Gorman Polar Form with possibly heterogeneous committed expenditures but a common price deflator; see Chipman and Moore (1990), McFadden (2004). In (21), this requires that the X functions be independent of income, so that discrete choices will exhibit no neoclassical income effect and HCV, HEV, and MCE coincide. In practical fulfillment of compensating transfers, the policymaker faces a decision-theory problem in which there will be social losses from under or over-compensation of individuals, and some (Bayesian) criterion must be applied to determine a loss-minimizing transfer rule. For example, if the policymaker has a quadratic social loss function, and a diffuse Bayesian prior, the optimal transfer to an individual equals the expected compensating transfer given observed characteristics. This suggests two rules in the case of partial observability. First, if transfers are fulfilled, prospectively or retrospectively, then they should equal the expected value of the exact compensating transfer given available information on the individual. Second, the impact of a policy change on a class of consumers in either prospective or retrospective applications should equal the expected value of the exact aggregate compensating transfers, with the appropriate compensating transfers determined by whether or not the transfers are hypothetical or fulfilled, and if the latter, whether this occurs before or after preference-maximizing choices in each scenario.

Relevant aggregates defined in Section 3 are the expected values UMCE(s,k) = $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|\mathrm{s,k}}\mathrm{MCE}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})$ and UMCE(s) = $\mathbf{E}_{\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon}|\mathrm{s}}\mathrm{MCE}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})$, or uniform Hicksian measures such as UHCV(s,k) and UHCV(s). Section 3 describes a computational approach to forming the relevant aggregates using a synthetic population; this approach can accommodate any assumptions the analyst chooses on the properties of $\boldsymbol{\varepsilon}$ and the observed histories on which the welfare measures are conditioned. However, in selected cases, it is possible to reduce computation by forming analytic expectations with respect to $\boldsymbol{\varepsilon}$. In the remainder of this section, we do this for the case where the scenario *a* choice is <u>not</u> observed, and three cases where this choice <u>is</u> observed: (A) all products, even "brands" whose prices and attributes do not change between scenarios, have distinct indices in the two scenarios; (B) "brands" with changing attributes and prices across scenarios have distinct indices, but benchmark "brands" whose attributes and prices do not change have the same indices; and (C) all "brands" are present in both scenarios *a* and *b*, and have the same indices, even though some have measured attributes or prices that change. In terms of choice sets, these cases are (A) $\mathbf{J}_a \cap \mathbf{J}_b = \emptyset$, (B) $\emptyset \neq \mathbf{J}_a \cap \mathbf{J}_{b\,a} \neq \mathbf{J}_a \cup \mathbf{J}_b$, and (C) $\mathbf{J}_a = \mathbf{J}_b$. Since the ε_j are approximation elements added for convenience, rather than utility components with deeper justification from consumer behavior, one should be able to pick from the cases (A)-(C) to get the most convenient computational formulas. However, if this makes a substantial difference in the overall level or distribution of compensating transfers, then (20) needs to be respecified to reduce the relative contribution of the ε_j elements. Case (A) is plausible if the consumer has a fixed idiosyncratic contribution ε_j to utility for each good j, but perceives of all goods in a new choice situation as if they were entirely new products. This stretches the neoclassical assumption of predetermined and fixed preferences, as it is equivalent to allowing a special preference tremble that can vary with choice situation. Cases (B) and (C) more easily fit the neoclassical interpretation of the ε_j as contributions from persistent unobserved attributes of branded products.

Consider the unconditional indirect utility function (22). Appendix B(b) shows that its expectation with respect to (β,σ,$\boldsymbol{\varepsilon}$), given history s, income I, and scenario m is

$$(29)\quad \mathbf{E}_{\beta,\sigma|s}\,\mathbf{E}_{\varepsilon}\max_{j\in\mathbf{J}_m}\,[I + \mathrm{v}(I,p_{jm},\mathbf{r}_m,z_{jm},\beta) + \sigma\varepsilon_j] = I + \mathbf{E}_{\beta,\sigma|s}\,\{\sigma\cdot\ln\textstyle\sum_{j\in J_m}\exp\big(\mathrm{v}_{jm}(I,\beta)/\sigma\big) + \sigma\cdot\gamma_0\},$$

where γ_0 denotes Euler's constant and $v_{jm}(I,\beta) \equiv v(I, p_{jm}, \mathbf{r}_m, z_{jm}, \beta)$. Scaling and differencing for m = *b*,*a*,

$$(30) \quad \text{UMCE}(s) \equiv \mathbf{E}_{k,\beta,\sigma,\boldsymbol{\varepsilon}|s}\text{MCE}(s,k,\beta,\sigma,\boldsymbol{\varepsilon}) = \mathbf{E}_{\beta,\sigma|s}[I_b - I_a + \sigma\cdot\ln\frac{\sum_{j\in J_b}\exp(v_{jb}(I_b,\beta)/\sigma)}{\sum_{j\in J_a}\exp(v_{ja}(I_a,\beta)/\sigma)}]/\mu(I_a,\beta,\sigma),$$

where the *mean scaling factor* $\mu(I_a,\beta,\sigma)$ is a weighted harmonic mean of the marginal utilities of income (26), with the MNL choice probabilities from (22) as weights, that depends on σ through these weights,

$$(31) \quad \frac{1}{\mu(I_a,\beta,\sigma)} \equiv \mathbf{E}_{k|,\beta,\sigma}\frac{1}{\mu_k(I_a,\beta)} = \sum_{k\in J_a}\frac{L_{km}(I_a,\beta,\sigma)}{\mu_k(I_a,\beta)}.$$

A Hicksian analogue of (30) is obtained by forming the expectation of (24) and solving

$$(32) \quad 0 = \mathbf{E}_{\beta,\sigma|s}[I_b - \text{UHCV}(s) - I_a + \sigma\cdot\ln\frac{\sum_{j\in J_b}\exp(v_{jb}(I_b-\text{UHCV}(s),\beta)/\sigma)}{\sum_{j\in J_a}\exp(v_{ja}(I_a,\beta)/\sigma)}].$$

A formula for UHEV(*s*) is more complicated. If scenario *a* optimal choices before and after the income transfer change due to income effects, then (29) is replaced by a complex expression from Appendix B(e). However, if the marginal utility of income $\mu_k(I_a,\beta)$ from (26) is independent of I_a and k remains the optimal choice after the transfer, then from (27), UHEV(*s*) = UMCE(*s*). Of course, if discrete choice exhibits no income effects, then the definitions above satisfy UMCE(*s*) = UHCV(*s*) = UHEV(*s*).

Next consider situations (A), (B), and (C) defined above in which choice of alternative k in scenario *a* is observed, $\delta_k(I_a) = 1$.

(A) From Appendix B(b), expected utility given the utility-maximizing choice k is the same as (29) when m = *a* and $I = I_a$. Further, independence implies that (29) will also apply when m = *b* and $I = I_b$. Then, (32) with conditioning on k added continues to hold in case (A), with a solution defining a uniform transfer UHCV(*s*,k) for each k. Further, (30) is altered only by substituting the choice-k specific scale factor (26), giving

$$(33) \quad \text{UMCE}(s,k) = \mathbf{E}_{\beta,\sigma|s,k}[I_b - I_a + \sigma\cdot\ln\frac{\sum_{j\in J_b}\exp(v_{jb}(I_b,\beta)/\sigma)}{\sum_{j\in J_a}\exp(v_{ja}(I_a,\beta)/\sigma)}]/\mu_k(I_a,\beta).$$

Note that (30) is the expectation of (33) with respect to the MNL probability $L_{km}(I_a,\beta,\sigma)$. While the formula (33) depends on k only due to the scale factor, its expectation conditioned on choice k in a population with heterogeneous observed environments will in general vary substantially with k due to selection on the environments that yield this choice.

(B) Suppose the alternatives in $\mathbf{J}_a \cup \mathbf{J}_b$ can be partitioned into a set A of alternatives with indices that appear only in $\mathbf{J}_a$; a set B of alternatives that appear only in $\mathbf{J}_b$; and a set C of "benchmark" alternatives appearing in both scenarios that satisfy $v_{ja}(I_a,\beta) = v_{jb}(I_b,\beta)$ for j ∈ C; this requires that their attributes and prices not change, and if these v_{jm} depend on income, that $I_a = I_b$. As a result of this assumption, the $v_{jm}(I,\beta) \equiv v_j(I,\beta)$ for j ∈ A∪B∪C do not depend on the scenario m.

Utilizing the conditional expectation formulas in Appendix B(c),

$$\text{(34)} \qquad \text{UMCE}(s,\text{k}) = \mathbf{E}_{\beta,\sigma|s,\text{k}}\left[I_b - I_a + \max_{j\in B\cup C}\left(v_j(I_b,\beta) + \sigma\varepsilon_j\right) - \left(v_k(I_a,\beta) + \sigma\varepsilon_k\right)\right]/\mu_k(I_a,\beta)$$

$$= \mathbf{E}_{\beta,\sigma|s,\text{k}}\frac{1}{\mu_k(I_a,\beta)}\left[I_b - I_a + \sigma\cdot\ln\left(\frac{\sum_{j\in B\cup C}\exp\left(v_j(I_b,\beta)/\sigma\right)}{\sum_{j\in A\cup C}\exp\left(v_j(I_a,\beta)/\sigma\right)}\right)\right] + \begin{cases}\mathbf{E}_{\beta,\sigma|s,\text{k}}\dfrac{L(\text{C}|\text{A,C})}{L(\text{A}|\text{A,C})}\,\dfrac{\sigma\cdot\ln(1-L(\text{A}|\text{A,B,C}))}{\mu_k(I_a,\beta)} & \text{if k} \in \text{A}\\ -\mathbf{E}_{\beta,\sigma|s,\text{k}}\dfrac{\sigma\cdot\ln(1-L(\text{A}|\text{A,B,C}))}{\mu_k(I_a,\beta)} & \text{if k} \in \text{C}\end{cases},$$

where $L(\text{A}|\text{A,B,C}) = \dfrac{\sum_{j\in A} e^{v_j(I_b,\beta)/\sigma}}{\sum_{j\in A\cup B\cup C} e^{v_j(I_a,\beta)/\sigma}}$ and $L(\text{C}|\text{A,C}) = \dfrac{\sum_{j\in C} e^{v_j(I_b,\beta)/\sigma}}{\sum_{j\in A\cup C} e^{v_j(I_a,\beta)/\sigma}}$.

The left-hand expectation term in the last line of (34) coincides with the expression (33) for UMCE(s,k) obtained when the idiosyncratic noise in scenario *b* is independent of the idiosyncratic noise in scenario *a*. The right-hand expectation term is an adjustment for the effect of the conditioning event on the expected maximum utility from B∪C, downward if k ∈ A and upward if k ∈ C. This expectation incorporates the effects of selection, which can be powerful if σ is large: Many choices from A will come from favorable draws of idiosyncratic noise even when observed attributes make these alternatives unattractive. Then, regression to the mean in draws of idiosyncratic noise will tend to make alternatives in B less desirable than their analogues in A even if they are objectively better. In contrast, when the analogues in B of alternatives in A objectively improve, choices from C that result from a favorable draw will lead to an even better expected outcome in scenario *b* since alternatives with this draw remain available. If the scale factors $\mu_k(I_a,\beta)$ vary with k, then the interaction of selection and income effects no longer gives the result that (30) with scale factor (31) equals the expectation of (34) with respect to k. For example, if alternatives in A have $\mu_k(I_a,\beta) > 1$, and alternatives in C have $\mu_k(I_a,\beta) = 1$, then the expectation of (34) with respect to k exceeds MCE(s) from (30). Equation (34) can also be adapted to calculate UHCV(s,k) for this case: Reduce income I_b in scenario b by UHCV(s,k), with this quantity adjusted so that (34) equals zero.

(C) Suppose $\mathbf{J}_a = \mathbf{J}_b = \mathbf{J} = \{0,\ldots,J\}$, $\delta_k(I_a) = 1$, and $\boldsymbol{\varepsilon}_a = \boldsymbol{\varepsilon}_b = \boldsymbol{\varepsilon}$, so that all alternatives are indexed the same and have the same idiosyncratic noise in both scenarios. It is possible to obtain analytic formulas for MCE(s,k) under quite general conditions in which the differences $v_{jb}(I_b^*, \beta) - v_{ja}(I_a^*, \beta)$, evaluated at income levels that may differ from I_a or I_b respectively due to transfers, vary across multiple alternatives. Appendix B(e) provides formulas that can be assembled to program this calculation, but these are too complex to be useful for comparison to the previous cases. We instead consider the special circumstance in which $I_a = I_b = I$ and the scenario affects only product J, so $v_{jm}(I,\beta)$ is independent of m for j < J. For this case, Appendix A(d) implies the following results:

If k = J, then

$$(35) \quad \text{UMCE}(s,\text{J}) = \mathbf{E}_{\beta,\sigma|s,\text{J}} \frac{1}{\mu_\text{J}(I,\beta)} \begin{cases} v_{\text{Jb}}(I,\beta) - v_{\text{Ja}}(I,\beta) & \text{if } v_{\text{Jb}}(I,\beta) > v_{\text{Ja}}(I,\beta) \\ \frac{\sigma}{L_{\text{Ja}}} \cdot \ln \frac{\sum_{j\in\text{J}} e^{v_{jb}(I,\beta)/\sigma}}{\sum_{j\in\text{J}} e^{v_{ja}(I,\beta)/\sigma}} & \text{if } v_{\text{Jb}}(I,\beta) < v_{\text{Ja}}(I,\beta) \end{cases},$$

while If k < J,

$$(36) \quad \text{UMCE}(s,\text{k}) = \mathbf{E}_{\beta,\sigma|s,\text{k}} \frac{1}{\mu_\text{k}(I,\beta)} \begin{cases} -\frac{L_{\text{Ja}}}{L_{\text{ka}}}(v_{\text{Jb}}(I,\beta) - v_{\text{Ja}}(I,\beta)) + \frac{\sigma}{L_{\text{ka}}} \cdot \ln \frac{\sum_{j\in\text{J}} e^{v_{jb}(I,\beta)/\sigma}}{\sum_{j\in\text{J}} e^{v_{ja}(I,\beta)/\sigma}} & \text{if } v_{\text{Jb}}(I,\beta) > v_{\text{Ja}}(I,\beta) \\ 0 & \text{if } v_{\text{Jb}}(I,\beta) < v_{\text{Ja}}(I,\beta) \end{cases},$$

where $L_{\text{ka}} = e^{v_{\text{ka}}(I,\beta)/\sigma} / \sum_{j\in\text{J}} e^{v_{ja}(I,\beta)/\sigma}$. As in case (B), this formula can be adapted to solve for the transfer UHCV(s,k) that when fulfilled makes a subsequent UMCE zero, while computation of UHEV(s,k) is in general more complicated.

We consider an example where due to a fixing agreement the price of a single product, say a tablet computer, is higher in scenario a than in scenario b. For the alternative configurations of $\mathbf{J}_a$, $\mathbf{J}_b$, and $\boldsymbol{\varepsilon}$, we estimate in Table 2 the measures $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,\text{k}}$MCE($s$,k,β,σ,**ε**), $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,\text{k}}$HCV($s$,k,β,σ,**ε**), and $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,\text{k}}$HEV($s$,k,β,σ,**ε**) in a synthetic population, and the measure UMCE(s,k). Suppose the product J = 1 has price p_{1m} in scenario m. Suppose the "no purchase" alternative has $p_{0m} = 0$. Suppose consumers have utilities of the form (21) with $v_{1m}(I, \beta) = \beta_1 I + \beta_2 - p_{1m}$ for alternatives where the product is purchased, and $v_{0m}(I, \beta) = 0$ for the "no purchase" alternative, for scenarios m = a,b. Then $\mu_0(I_a,\beta) = 1$ and $\mu_1(I_a,\beta) = 1 + \beta_1$. The idiosyncratic noise cases we consider are (A) independent noise across scenarios, represented by J_a = (0,1} and J_b = {2,3}, with j = 0,2 corresponding to "no purchase" and j = 1,3 corresponding to "purchase"; (B) J_a = {0,1} and J_b = {0,3}, with j = 0 corresponding to a common "no purchase" and j = 1,3 corresponding to "purchase"; and (C) $J_a = J_b$ = {0,1}, with j = 0 corresponding to a common "no purchase" and j = 1 to a common purchase. Suppose that $\beta_1 = 0.002$ and $\sigma = 9$ are fixed parameters, and that β_2 is normal with mean zero and standard deviation 60. The choice probabilities are then mixed logit, with $P_{0m}(I) = \mathbf{E}_\beta \frac{1}{1+\exp(v_{jm}(I,\beta)/\sigma)}$ for non-purchase of the product j in scenario m. Suppose the consumer faces p_{1a} = \$110 and p_{1b} = \$90, and the base income is I = \$50,000. The probabilities of buying the product in a synthetic population of 10,000 are $P_{1a}(50000) = 0.430$, $P_{1b}(50000) = 0.555$, and $P_{1a}(56000) = 0.505$. These probabilities imply an arc income elasticity of 1.45 and an arc price elasticity of -1.59 for the given market changes. The table shows first that for this example, $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,s,k,}$MCE($s$,k,β,σ,$\boldsymbol{\varepsilon}$), $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,s,k,}$HCV($s$,k,β,σ,$\boldsymbol{\varepsilon}$), and $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,k}$HEV($s$,k,β,σ,$\boldsymbol{\varepsilon}$) estimated in the synthetic population are almost the same. This result is consistent with the conclusion of Willig (1976) that income effects are typically small. The value of UMCE using an analytic expectation with respect to ε differs modestly from the synthetic population estimate of $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,s,k,}$MCE($s$,k,β,σ,$\boldsymbol{\varepsilon}$), but the difference is well within sampling error. The Marshallian consumer surplus, estimated here using the trapezoid rule, is nearly identical to UMCE.

Table 2. Comparisons of Welfare Measures (\$pp)

	Case A	**Case B**	**Case C**	**σ = 0**
Total Population				
Marshallian consumer surplus	9.819	9.799	9.781	9.854
UMCE (analytic $\mathbf{E}_\varepsilon$)	9.848	9.858	9.802	9.840
MCE (synthetic population)	9.886	9.806	9.755	9.840
HCV (synthetic population)	9.883	9.803	9.753	9.837
HEV (synthetic population)	9.886	9.806	NC	9.840
Class of Product Purchasers				
UMCE (analytic $\mathbf{E}_\varepsilon$)	18.568	18.283	19.960	19.960
MCE (synthetic population)	18.609	18.368	19.960	19.960
HCV (synthetic population)	18.610	18.368	19.960	19.960
HEV (synthetic population)	18.609	18.368	NC	19.960
Class of Non-Purchasers				
UMCE (analytic $\mathbf{E}_\varepsilon$)	3.305	3.535	2.180	2.240
MCE (synthetic population)	3.340	3.382	2.097	2.240
HCV (synthetic population)	3.335	3.375	2.093	2.235
HEV (synthetic population)	3.340	3.382	NC	2.240

The example suggests that UMCE will be an adequate approximation when $\mathbf{E}_{\beta,\sigma,\boldsymbol{\varepsilon}|s,k_a}$HCV($s$,k,β,σ,$\boldsymbol{\varepsilon}$) is the ideal measure. However, the closeness of UMCE and expected HCV is sensitive to the magnitude of the change in price in the two scenarios, and larger changes can lead to a gap between these measures. In short, when UMCE is used as an approximation to expected HCV, it is desirable to use synthetic population methods with large samples to check the quality of the approximation.

There is variation in the welfare measures when one moves from Case A with independent disturbances to Case C with common disturbances. In particular, Cases A and B attribute less welfare gain to purchasers and more welfare gain to non-purchasers than does Case C and a σ = 0 case with no idiosyncratic noise. Then, the assumptions made on the persistence of idiosyncratic errors across scenarios makes a difference.

Table 3. Effect of Idiosyncratic Noise on Distribution of Welfare Changes ($pp)

	Case A			***Case B***			**Case C**		
UMCE at	Buyers	Non-buyers	Total	Buyers	Non-buyers	Total	Buyers	Non-buyers	Total
σ = 0	19.960	2.240	9.840	19.960	2.240	9.840	19.960	2.240	9.840
σ = 9	18.568	3.305	9.848	18.283	3.535	9.858	19.960	2.180	9.802
σ = 36	13.641	6.943	9.895	5.046	14.057	10.086	19.960	1.683	9.737
σ = 64	11.687	8.463	9.927	-9.086	25.560	10.370	19.960	1.224	9.734

Table 3 continues the example with different scale factors σ, and shows that at high levels of σ relative to the observed changes in the scenarios, the effects of selection on idiosyncratic noise can drastically alter the distribution of welfare gains between purchasers and non-purchasers. We infer from this table that unless there is compelling evidence to support the case (B) assumptions, they should be rejected in favor of case (A) or case (C) assumptions that more closely approximate a model in which neoclassical tastes, heterogeneous across consumers but durable within each consumer, describe choice behavior without significant added idiosyncratic noise. Finally, there is a substantial advantage in simplicity for the analytic expectations coming out of the case (A) compared to case (C), suggesting that case (C) be used only if there is persuasive evidence for durable idiosyncratic noise.

6. DECISION-UTILITY VERSUS EXPERIENCED-UTILITY

Decision-utility is defined as the objective function that the neoclassical consumer optimizes in making her market choices, the function that can be recovered (with money-metric scaling) from sufficiently rich observations on these market choices. The foundations of welfare theory restated in Section 3 assume that decision-utility is a direct and complete measure of well-being. In reality, anticipated decision-utility and realized experienced-utility can differ. The most straightforward case, covered by neoclassical theory, is decision-making under uncertainty where the decision utility function equals the expectation over objective probabilities of a utility function of outcomes, and the experienced utility function equals this utility function evaluated at the realized outcome. For example, consumers may be uncertain about attributes of alternatives such as product durability, so that buying a product is equivalent to buying a lottery ticket on its attributes. Under von Neumann-

Morgenstern assumptions on utility, sufficiently rich market observations on choice among risky prospects will suffice to recover the utility function of outcomes.

Beyond neoclassical decision-making under uncertainty, there are a number of factors that can cause gaps between decision-utility and experienced-utility: (1) misperceptions of shrouded, ambiguous, or misleadingly promoted product attributes, (2) unrealistic personal probability judgements on uncertain events, (3) whims and psychometric noise that induce tremble in tastes, as in Thurstone (1927), (4) factors that influence the sensation of well-being but do not influence market choices, such as provision of pure public goods and services, (5) inconsistencies in preferences, such as time-inconsistent discounting and unanticipated habit-formation or addiction, and (6) flaws in the process of utility maximization, such as reference point bias and hypersensitivity to recent experience.

When there are gaps between decision-utility and experienced-utility, which should be used to measure well-being? Roughly, welfare measures based on decision-utility focus on equity in opportunity, while those based on experienced-utility focus on equity in outcomes. It may seem evident that consumer perceptions and decision-making in markets are simply instruments to achieve final outcomes, so experienced utility should be at the core of welfare assessment. However, there are complicating factors. First, while decision-utility is arguably linked to and recoverable from observed market behavior of consumers, there will often be no clear link between decision-utility and experienced-utility, and no reliable method of recovering experienced-utility from economic or psychological experiments. Then, it may be impossible to use experienced utility as a basis for evaluating transfers or other policies designed to address inequitable outcomes. Second, when consumers are fully and accurately informed about the prospects and contingencies they face, and there are sufficient contingent markets so that they can insure against risks if they choose, then they have it in their own hands to make informed choices and live with the consequences of these choices. Further, interventions based on experienced utility can introduce "moral hazard" in which the anticipation of ex post remedies for bad outcomes leads consumers to take excess risks and be less diligent in their decisions, particularly by failing to take steps to avoid or mitigate harm. Then, for fully informed consumers facing complete contingent markets, policies should arguably be evaluated in terms of decision-utilities. However, when consumers are poorly informed or lack opportunities to manage risks, ex post equity is a social concern, and/or consumers are unable to look after their own interests, interventions by a benevolently paternalistic regulator may be appropriate, with or without a basis in experienced utility.

In general, it will be important to know how perceptions, decision utility, and experienced utility are linked. Misperceptions of attributes and biased personal probabilities, listed above as sources (1) and (2) of gaps between decisions and experience, do not substantially alter the neoclassical preference structure, and can in principle be accounted for starting from decision utility and correcting these factors. For example, it should be straightforward in principle to correct consumer misperceptions arising from supplier misrepresentation of product attributes. In practice, identification and recovery of personal perceptions and probabilities may overburden market data and require extra-market experimental observations.

Instabilities in tastes arising from psychometric noise, source (3) above, also leaves many neoclassical elements of choice in place. However, preference tremble creates a fundamental difficulty with welfare analysis: How to measure welfare changes when preferences are not fixed. One tack is to simply take the preferences revealed in "as is" decisions as yardsticks for welfare comparisons, and ignore shifts in "but for" tastes caused by tremble. An issue here is that incorporating whims into the welfare calculus can make the results sensitive to selection effects, as in the previous section. Another tack is to try to recover stable "core" preferences stripped of the tremble introduced by transient whims and misperceptions. However, a measured distribution of preferences that is a convolution of population heterogeneity and individual tremble will confound recovery of either component of the convolution, making recovery of core preferences problematic unless one observes multiple choices for each individual.

Other elements entering experienced utility such as sources (4)-(6) of gaps between perceptions and experience, particularly factors that leave no trace in market choices, can confound choice behavior so that there may be no identifiable decision-utility or linked experienced utility that capture consumer well-being. Then recovery of experienced-utility will be beyond the capacity of economists using customary market-based data. Ben-Akiva, McFadden, and Train (2016) discuss experimental methods for direct elicitation of preference that might be used in principle to address these identification and recovery tasks. Stated preference methods are widely used in market research to forecast demand for new products and the value of extra-market resources, with varying degrees of reliability; see McFadden (2017). An open question about measurement of well-being of consumers who have behavioral elements in their decision-making and a gap between decision-utility and experienced-utility is whether experienced utility could be elicited directly in conjoint analysis experiments, either through experiments used to uncover the components of experienced utility, or through conjoint elicitation methods such as elicitation of stated personal probabilities. Established experimental designs for such elicitations are not available now, and there are major scientific challenges to their development, particularly known biases in personal probability judgements and the problem of verification, but there will be high payoffs to future scientific breakthroughs in these areas. Since the focus of this paper is welfare analysis using market observations, this paper will not explore experimental, cognitive, or neurological approaches to direct measurement of well-being.

To facilitate analysis of the consequences of gaps between anticipation and experience, let superscript "d" denote decision utility and superscript "e" denote experienced utility. From (20), choice in scenario m at income I then maximizes $v^{d}_{jm}(I,\beta^{d}) + \sigma\varepsilon_j$; let $\delta_{jm}(I,\beta^{d},\sigma,\boldsymbol{\varepsilon})$ denote an indicator for this choice, and $P_{jm}(I,\beta^{d},\sigma)$ the probability of this choice given β^d,σ. Let $v^{e}_{jm}(I,\beta^{d}) + \sigma\varepsilon_j$ denote the experienced utility obtained from choice j. The application will determine the structure of $v^{e}_{jm}(I,\beta^{e})$ and its linkage to $v^{d}_{jm}(I,\beta^{d})$, and the mapping from β^d to β^e.[17] The decision utility and experienced utility from a choice situation with income I are, in this notation,

$$(37) \quad u^{d}_{m}(I,\beta^{d},\sigma) = I + \max_{j=0,\ldots,J_m}\left[v^{d}_{jm}(I,\beta^{d}) + \sigma\varepsilon_j\right]$$

and

$$(38) \quad u^{e}_{m}(I,\beta^{e},\beta^{d},\sigma) = I + \sum_{j=0}^{J_m}\delta_{jm}(I,\beta^{d},\sigma,\boldsymbol{\varepsilon})\left[v^{e}_{jm}(I,\beta^{e}) + \sigma\varepsilon_j\right]$$
$$\equiv u^{d}_{m}(I,\beta^{d},\sigma) + \sum_{j=0}^{J_m}\delta_{jm}(I,\beta^{d},\sigma,\boldsymbol{\varepsilon})\left[v^{e}_{jm}(I,\beta^{e}) - v^{d}_{jm}(I,\beta^{d})\right],$$

where experienced utility in the last expression equals anticipated utility plus a correction that comes from differences in anticipated and realized attributes and tastes. Combined with (25) defining MCE^d for decision-utility, (38) implies the experienced-utility welfare measure, given (β^d,β^e,σ), and $\delta_{ka}(I,\beta^{d},\sigma,\boldsymbol{\varepsilon}) = 1 = \delta_{jb}(I,\beta^{d},\sigma,\boldsymbol{\varepsilon})$,

$$(39) \quad \mu^{e}_{k}(I_a,\beta)\cdot \mathrm{MCE}^{e}(s,k,\beta^{e},\sigma^{e},\boldsymbol{\varepsilon}^{e}_{\mathbf{a}},\boldsymbol{\varepsilon}^{e}_{\mathbf{b}},\beta^{d},\sigma^{d},\boldsymbol{\varepsilon}^{d}_{\mathbf{a}},\boldsymbol{\varepsilon}^{d}_{b}) = \mu^{d}_{k}(I_a,\beta)\cdot \mathrm{MCE}^{d}(s,k,\beta^{d},\sigma^{d},\boldsymbol{\varepsilon}^{d}_{\mathbf{a}},\boldsymbol{\varepsilon}^{d}_{b})$$
$$+ v^{e}_{jb}(I_b,\beta^{e}) - v^{d}_{jb}(I_b,\beta^{d}) - v^{e}_{ka}(I_a,\beta^{e}) + v^{d}_{ka}(I_a,\beta^{d}),$$

where $\mu^{e}_{k}(I_a,\beta)$ is the marginal experienced utility of income.

Economists should be very cautious in applying the traditional welfare calculus when decision-utility requires behavioral factors to explain behavior; as transfers to maintain decision utility can have unreliable and unintended effects on experienced well-being. If anticipated tastes are an unreliable guide to realized tastes, this is a challenge to the foundations of welfare economics; see Lowenstein and Ubel (2008), Thaler and Sunstein (2003,2008),

[17] One case which is straightforward occurs when $v^{d}_{jm}(I)$ and $v^{e}_{jm}(I)$ differ only because of differences in observed anticipated and experienced product attributes, z^{d}_{jm} and z^{e}_{jm}, due to say false advertising of attributes, and $\beta^d = \beta^e$. Cases that are more challenging for economic analysis occur when either anticipated or experienced attributes are unobserved, or v^e and v^d differ due to optimization errors and volatility in tastes. In such cases, the analyst will often have no recourse other than using extra-market observations such as experimental elicitation of stated preferences, with attendant questions of reliability.

McFadden (2014), Train (2015), Bernheim (2016). There is currently no accepted general welfare theory for non-neoclassical consumers who have shifts between anticipated and realized tastes, even though the random decision-utility setup itself can accommodate many non-neoclassical elements. However, there may be some special circumstances and assumptions that overcome this limitation. For example, differences in "as is" or "but for" $(z_{\mathrm{jm}}^{\mathrm{d}}, p_{\mathrm{jm}}^{\mathrm{d}})$ and $(z_{\mathrm{jm}}^{\mathrm{e}}, p_{\mathrm{jm}}^{\mathrm{e}})$ may be limited to identifiable misperceptions such as misinformation about product attributes, and the joint distribution of anticipated and realized tastes may by assumption be generated through limited differences such as personal misjudgments on the probabilities of contingent events or biases in risk preferences and time discounts used in making decisions. If it is plausible that such limited shifts in tastes can be fully described and modeled using specific external evidence, then welfare analysis based on (39) may be justified.

An example of consumer behavior that appears to be distorted by unrealistic personal probability judgements is consumer choice of health insurance policies. An argument, simplified from Heiss, McFadden, and Winter (2013) and McFadden and Zhou (2015), shows that misperceptions can be identified and corrected in some cases. Suppose consumers face stochastic medical expenses *c*, and have the subjective perception that these have a distribution $\mathrm{K}^{\mathrm{d}}(c)$ with a mean μ_{d} and variance $\mathrm{s_d}^2$. Suppose they have a menu of insurance alternatives j = 0,…,J with plan j characterized by a premium $\mathrm{p_j}$ and a copayment rate $\mathrm{r_j}$, with $\mathrm{p_0} = 0$ and $\mathrm{r_0} = 1$. Suppose their decision-utility is a money-metric transformation of a constant-absolute-risk-aversion (CARA) expected utility function,

$$(40) \qquad \mathrm{u_j} = \frac{-1}{\beta}\ln\int_{c=0}^{+\infty}\exp\left(-\beta\left(I - p_{\mathrm{j}} - \mathrm{r_j}c\right)\right)\mathrm{K^d}(\mathrm{d}c) + \sigma\varepsilon_{\mathrm{j}} \equiv I - p_{\mathrm{j}} - \kappa^{\mathrm{d}}(\beta\mathrm{r_j})/\beta + \sigma\varepsilon_{\mathrm{j}},$$

where *I* is income, κ^{d} is the cumulant generating function of $\mathrm{K^d}$, β is a risk-aversion parameter with a probability distribution in the population, and the parameter σ scales psychometric noise ε_{j}. Replacing the cumulant generating function κ^{d} in (40) with a quadratic approximation gives a utility $\mathrm{u_j} = I - p_{\mathrm{j}} - \mu_{\mathrm{d}}\mathrm{r_j} - \frac{1}{2}\mathrm{s_d^2}\beta\mathrm{r_j^2} + \sigma\varepsilon_{\mathrm{j}}$ of the form (21). Suppose $(\ln\beta, \ln\sigma)$ is distributed bivariate normal, and the ε_{j} is i.i.d. EV1. Then observations on consumer insurance choices in real or experimental markets allows estimation of the parameters of the bivariate normal distribution, and μ_{d}, and $\mathrm{s_d}^2$. Observations on <u>objective</u> probabilities $\mathrm{K^e}(c)$ for health expenses allow estimation of μ_{e} and $\mathrm{s_e}^2$. Then specialization of (40) using (41) and the quadratic approximations to the cumulant generating functions κ^{d} and κ^{e} allow estimation of the money-metric loss in consumer utility arising from poor choices due to misperception of medical expense risk.

7. WELFARE CALCULUS FOR COMMON POLICY PROBLEMS

Suppose mixed MNL choice probabilities of the form (22), along with the associated parameter α of a population distribution of taste parameters $F(\beta,\sigma|\alpha)$ and a money-metric utility of the form (21), have been estimated from choice data collected in real or hypothetical markets. Using these estimates, *prospective* benefit-cost analysis using decision utility can be carried out by solving (24) or evaluating (25) for each consumer in a synthetic population defined by draws of s, parameters (β,σ) from $F(\beta,\sigma|s,\alpha)$, and idiosyncratic noise $\boldsymbol{\varepsilon}$. Measures such as HCV, HEV, or MCE can then be averaged over the synthetic consumers falling into classes defined by restrictions on s, with the law of large numbers operating to ensure reliable estimates of the net transfer to the class that when optimally distributed leaves its members indifferent to the policy change. Alternately, one can concentrate on estimating a UMCE measure (30) for this class. To simplify notation, suppress the “d” superscript for decision utility. Let **C** denote the set of alternatives whose attributes are unchanged by a shift from policy *a* to policy *b*. By construction, **C** is a proper subset of $\mathbf{J}_a$ and $\mathbf{J}_b$ which contain alternatives whose attributes do not change, and **C** always contains at least j = 0. Then (30) can be rewritten as

$$(41)\qquad \text{UMCE}(s) \equiv \mathbf{E}_{\beta,\sigma|s}\left[I_b - I_a + \sigma\cdot\ln\frac{L_{\mathbf{C}a}(I_a,\beta,\sigma)}{L_{\mathbf{C}b}(I_b,\beta,\sigma)}\right]/\mu(I_a,\beta,\sigma)$$

where $L_{\mathbf{C}m}$ is the logit probability at random parameters (β,σ) of choice from **C** in scenario m. For example, introducing a set **B** of new products with attributes included in $(\mathbf{x}_b,\mathbf{p}_b)$, keeping unchanged the attributes of existing products in **C**, is $\text{UMCE}_b(s) = \mathbf{E}_{\beta,\sigma|s}[I_b - I_a - \sigma\cdot\ln L_{\mathbf{C}b}(I_b,\beta,\sigma)]/\mu(I_a,\beta,\sigma)$.

For small policy changes and $\mathbf{J}_a = \mathbf{J}_b = \mathbf{J}$, a Taylor’s expansion of the first form of (41) in variations $\Delta v_j \equiv v_{jb}(I_b,\beta) - v_{ja}(I_a,\beta) \equiv \Delta X_j\beta - \Delta p_j$, where $\Delta X_j \equiv X(I_b - p_{jb},\mathbf{r}_b,z_{jb}) - X(I_a - p_{ja},\mathbf{r}_a,z_{ja})$ and $\Delta p_j \equiv p_{jb} - p_{ja}$, gives the approximation,

$$(42)\qquad \text{UMCE}(s) = \mathbf{E}_{\beta,\sigma|s}[I_b - I_a + \textstyle\sum_{j\in\mathbf{J}}[L_{ja}(I_a,\beta,\sigma)\Delta v_j + O\big((\Delta v_j)^2/\sigma\big)]]/\mu(I_a,\beta,\sigma)\ .$$

Another useful approximation, due to Doug MacNair, applies the expansion $\ln(1 - y) = -y + O(y^2)$ to (41) with $L_{\mathbf{B}b} = 1 - L_{\mathbf{C}b}$ and $L_{\mathbf{A}a} = 1 - L_{\mathbf{C}a}$ the probabilities of choosing products whose attributes change, to obtain

$$(43)\qquad \text{UMCE}_b(s) = \mathbf{E}_{\beta,\sigma|s}[I_b - I_a + \sigma\cdot[L_{\mathbf{B}b}(I_b,\beta,\sigma) - L_{\mathbf{A}a}(I_a,\beta,\sigma) + O((1 - L_{\mathbf{C}m})^2]]/\mu(I_a,\beta,\sigma).$$

When σ and $\mu = \mu(I_a,\beta,\sigma)$ are homogeneous in the population, (43) has a particularly simple form,

$$(44)\qquad \text{UMCE}_b(s) \approx [\frac{I_b - I_a}{\mu} + \frac{\sigma}{\mu}\mathbf{E}_{\beta|s}[L_{\mathbf{B}b}(I_b,\beta,\sigma) - L_{\mathbf{A}a}(I_a,\beta,\sigma)]],$$

the income difference scaled by μ plus the difference in the (σ/μ)-scaled full population market share of consumers choosing the products affected by the policy change. For example, if a set **A** of products is "new" in scenario *b*, and income and the attributes of the remaining products in **C** are unchanged, then $\mathrm{UMCE}_b(s) \approx \frac{\sigma}{\mu}\cdot\mathbf{E}_{\beta|s}L_{\mathbf{A}b}(I_b,\beta,\sigma)]$, a scaled market share of the new products.

Consider a policy that affects attributes and price of products, and let $X_{j\lambda} = X_{ja} + \lambda\Delta X_j$, $p_{j\lambda} = p_{ja} + \lambda\Delta p_j$, and $I_\lambda = I_a + \lambda(I_b - I_a)$ for $\lambda \in [0,1]$ and j = 1,…,J denote a linear path that achieves this change. Let UMCE_λ denote (41) evaluated at point λ on this path. Let $v_{j\lambda} \equiv X_{j\lambda}\beta - p_{j\lambda}$ and $L_{j\lambda} = e^{v_{j\lambda}/\sigma}/\sum_{i\in J} e^{v_{i\lambda}/\sigma}$. Then $\partial v_{j\lambda}/\partial\lambda = \Delta X_j\beta - \Delta p_j$, and since the numerator of $L_{C\lambda}$ does not vary with λ,

$$(45)\qquad \frac{d\,\mathrm{UMCE}_\lambda(s)}{d\lambda} = \mathbf{E}_{\beta,\sigma|s}[I_b - I_a + \textstyle\sum_{j=1}^{J} L_{j\lambda}(\Delta X_j\beta - \Delta p_j)]/\mu(I_a,\beta,\sigma).$$

Then the incremental change in UMCE is a demand-weighted average of the changes $\Delta X_j\beta - \Delta p_j$ in the systematic components of utility. First, consider the common circumstance where ΔX_j and Δp_j do not depend on *s*; this will be the case for example for a product offered in a national market where interactions of product attributes and individual characteristics are not needed to explain choice behavior. Then, (45) reduces to

$$(46)\qquad \frac{d\,\mathrm{UMCE}_\lambda(s)}{d\lambda} = \mathbf{E}_{\beta,\sigma|s}[I_b - I_a + \textstyle\sum_{j=1}^{J}[\Delta X_j\bar{\beta}_{j\lambda} - \Delta p_j]L_{j\lambda}(I_\lambda,\beta,\sigma)/\mu(I_a,\beta,\sigma),$$

where $\bar{\beta}_{j\lambda} = \frac{\mathbf{E}_{\beta,\sigma|s,\alpha}\beta L_{j\lambda}(I_\lambda,\beta,\sigma)}{\mathbf{E}_{\beta,\sigma|s,\alpha}P_j(L_{j\lambda},\beta,\sigma)}$ denotes the mean of β *among consumers who choose* j when the alternatives are characterized by $(\mathbf{X}_\lambda,\mathbf{p}_\lambda)$. In this case, $\bar{\beta}_{j\lambda}$ gives WTP for attribute changes that translate directly into incremental compensating variation. In the special sub-case of changes that are uniform in j, $\Delta x_j = \Delta x_1$ for $j \neq 0$, $\bar{\beta}_{j\lambda}$ is independent of j and is the mean of β among all buyers. In the special case that the relevant components of β are homogeneous, then $\bar{\beta}_{j\lambda} = \beta$ in the corresponding components, and these coefficients are unequivocal measures of "part-worths". More generally, obtaining $\bar{\beta}_{j\lambda}$ is a calculation that requires estimates of both F(β,σ|*s*,α) and $L_{j\lambda}(I_\lambda,\beta,\sigma)$.

Second, when the relevant components of β are homogeneous, but Δx_j and Δp_j are heterogeneous over *s*, (45) reduces to

$$(47)\qquad \frac{d\,\mathrm{UMCE}_\lambda(s)}{d\lambda} = \mathbf{E}_{\beta,\sigma|s}\left[I_b - I_a + \left(\overline{\Delta X_j}\right)_\lambda\cdot\beta - \left(\overline{\Delta p_j}\right)_\lambda\right]/\mu(I_a,\beta,\sigma),$$

so the relevant components of β give WTP for mean attribute changes among consumers choosing j when product features are described by $(\mathbf{X}_\lambda,\mathbf{p}_\lambda)$. Third, when β is heterogeneous and the ΔX_j are heterogeneous over the population (i.e., vary with s), the relationship between values of β and $\mathbf{E}_{\beta,\sigma|s,\alpha}\mathrm{UMCE_b}(\beta,\sigma)$ is more complex; (42) requires a calculation that handles selection driven by both consumer history and taste heterogeneity.

The scaling parameter σ appears in (41) and (43) to have a prominent direct role in determining the level of UMCE(s), but (42) indicates that this is offset elsewhere, so that the final impact of σ is only indirect, through its influence on the choice probabilities. To see this more generally, write (41) as

$$(48)\quad \mathrm{UMCE}(s) \equiv \int_{\lambda=0}^{1} \frac{\mathrm{dUMCE}_\lambda(s)}{\mathrm{d}\lambda}\,\mathrm{d}\lambda \;=\; \mathbf{E}_{\beta,\sigma|s}[I_\mathrm{b} - I_\mathrm{a} + \textstyle\sum_{j\in\mathbf{J}}\int_{\lambda=0}^{1} L_{j\lambda}\big(\Delta\mathrm{X}_j\beta - \Delta p_j\big)\mathrm{d}\lambda]/\mu(I_\mathrm{a},\beta,\sigma).$$

This is a line integral over the rectifiable path of the area behind the demand functions for the products in **A** between the old and new vectors of quality-adjusted net values, which is the Mashallian consumer surplus associated with the change from policies *a* to *b*; (48) depends on σ only through its influence on the acuity of consumer response to price changes. These price effects are usually bounded even when there is a positive probability of very small σ. Recall that the own price elasticity of a MNL probability $L_{j\lambda} = \frac{\exp(\frac{v_{j\lambda}}{\sigma})}{\sum_{i\in\mathbf{J}}\exp(\frac{v_{i\lambda}}{\sigma})}$ equals $\frac{-p_{j\lambda}(1-L_{j\lambda})}{\sigma}$. Use the inequality $e^{-c/\sigma} \le \sigma/c$ for $c > 0$. If $c_k = \max_{j=0,\dots,J}(v_{j\lambda} - v_{k\lambda}) > 0$, then $L_{k\lambda}$ is bounded above by σ/c_k, and if $c_{-k} \equiv \min_{i\neq k} c_i > 0$, then $1 - L_{k\lambda}$ is bounded by $J\sigma/c_{-k}$. Then the price elasticity is bounded in magnitude by $\max\{p_{k\lambda}/c_k, Jp_{k\lambda}/c_{-k}\}$ no matter how small σ. The limited sensitivity of (48) to σ is also seen by considering limiting cases. For constant $\sigma \longrightarrow 0$, $\mathrm{UMCE}(s) \longrightarrow \mathbf{E}_{\beta,\sigma|s}\,\{\min_{j\in\mathbf{J}} v_{j\lambda} - \min_{j\in\mathbf{J}} v_{ja}\}/\mu(I_a,\beta,\sigma)$, and for $\sigma \longrightarrow +\infty$, $\mathrm{UMCE}(s) \longrightarrow \mathbf{E}_{\beta,\sigma|s}\,\frac{1}{|\mathbf{J}|}\sum_{j\in\mathbf{J}}\{v_{j\lambda} - v_{ja}\}/\mu(I_a,\beta,\sigma)$. The difference in these expressions comes only from the difference between least and average quality-adjusted net values, reflecting two extremes in the acuity of consumers in gravitating to alternatives with the greatest quality-adjusted net values.

Next consider *retrospective* welfare analysis that quantifies the harm to consumers from a past “as is” scenario *a* compared to a “but-for” scenario *b* in which product attributes are changed by altering attributes or seller conduct judged defective or improper. By its nature, retrospective analysis deals with loss of experienced utility, and with compensating transfers to make consumers whole after their “as is” choices have been made, so that experienced-utility MCE rather than HCV or HEV is the target of the analysis, even if choices are influenced by neoclassical income effects. The analysis in these applications is focused on objective changes in product attributes rather than shifts in consumer tastes, so it is reasonable to assume that the decision-utility and experienced-utility tastes are the same, and that in most cases $I_a = I_b$, so that any gap between decision utility and experienced utility comes from differences in z_{jm}^{d} and z_{jm}^{e}. Then experienced utility MCE is given by (39), with

MCE^{d} given by (25). The circumstances of the application will determine the configurations of $v_{jm}^{d}(I, \beta, \sigma)$ and $v_{jm}^{e}(I, \beta, \sigma)$ that prevail. A critical question is whether consumers are fully and accurately informed about the attributes of products in both the "as is" and "but for" scenarios, or whether the issue is misinformation or deception on product attributes in the "as is" scenario.

The first case we consider is one in which consumers have full information on the available products under both "as is" and "but for" conditions. One example is anti-trust litigation in which the question is the harm to consumers caused by improper supplier conduct such as price collusion, market allocation, bundling, or artificial barriers to entry. Other examples are environmental litigation in which the question is the harm caused by improper disposal of hazardous wastes, and patent litigation in which the question is the value to consumers of infringing features. With full information, anticipations are realized, so that $v_{jm}^{d}(I, \beta) = v_{jm}^{e}(I, \beta)$ for m = *a*,*b*. In many applications, the class of consumers of interest is not the general population, but individuals meeting specific conditions, such as residence in a specified region. If the class is defined by consumer characteristics *s* in a set **T**, and either the "as is" choice is unobserved, or it is observed but $\mathbf{J}_a \cap \mathbf{J}_b = \emptyset$, then the per capita transfer prescribed for this class is

$$(49) \qquad \mathrm{UMCE}(\mathbf{T}) = \mathbf{E}_{s|\mathbf{T}} \mathbf{E}_{\beta,\sigma|s} \sigma \cdot \ln \frac{L_{Ca}(I_a, \beta, \sigma)}{L_{Cb}(I_b, \beta, \sigma)} / \mu(I_a, \beta, \sigma).$$

Next consider cases where consumers are misinformed about products in the "as is" scenario 1, due to failure to deliver goods as promised, or to deceptive advertising, resulting in experienced utility that deviates from anticipated utility, an application studied by Chorus and Timmermans (2009), Alcott (2013), Schmeiser (2014), and Train (2015). In these cases, consumers are fully informed in scenario *b*. Then, $\mathbf{J} = \mathbf{J}_a = \mathbf{J}_b$, (z_{ja}^{d}, p_{ja}^{d}) and (z_{ja}^{e}, p_{ja}^{e}) are different for j in a set of products **D** where the misinformation occurs in scenario *a*, but (z_{ja}^{d}, p_{ja}^{d}) and (z_{ja}^{e}, p_{ja}^{e}) agree for $j \notin \mathbf{D}$ and (z_{jb}^{d}, p_{jb}^{d}) and (z_{jb}^{e}, p_{jb}^{e}) agree for all j. Continue to assume that anticipated and experienced taste parameters are the same. There are two leading possibilities for defining the "but for" scenario: the benchmark "but for" net values can match either the anticipated decision-utility net values when the anticipation is accurate, or match the realized utility net values when these net values are correctly anticipated in the "as is" scenario. The former benchmark applies to contract violations, where the violator is obligated to provide the promised product or equivalent compensation. The latter benchmark arguably applies to false advertising cases where the "but for" scenario correctly informs consumers of the actual product attributes, so that anticipations are realistic.

In the contract violation case, the "but for" net values are defined to match what consumers *anticipated* in the "as is" situation. Then $v^{e}_{jb}(I,\beta)$ = $v^{d}_{jb}(I,\beta)$ = $v^{d}_{ja}(I,\beta)$ for all j, but $v^{e}_{ja}(I,\beta) \neq v^{d}_{ja}(I,\beta)$ for j ∈ **D**. The appropriate metric for comparing consumer welfare under the "as is" and the "but for" scenarios for consumers with the observed "as is" choice k is the experienced MCE, which becomes

$$(50)\quad \mathrm{MCE^{e}}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon}) = \left\{\max_{j\in\mathbf{J}}\left[v^{d}_{jb}(I_b,\beta) + \sigma\varepsilon_{jb}\right] - \left[v^{e}_{ka}(I_a,\beta) + \sigma\varepsilon_{ka}\right]\right\}/\mu_k(I_a,\beta)$$

$$= \{\max_{j\in\mathbf{J}}\left[v^{d}_{jb}(I_b,\beta) + \sigma\varepsilon_{kb}\right] - \max_{j\in\mathbf{J}}\left[v^{d}_{ja}(I_a,\beta) + \sigma\varepsilon_{ja}\right] + v^{d}_{ka}(I_a,\beta) - v^{e}_{ka}(I_a,\beta)\}/\mu_k(I_a,\beta).$$

If $\varepsilon_a = \varepsilon_b$, this expression reduces to $\mathrm{MCE^{e}}(s,\mathrm{k},\beta,\sigma,\boldsymbol{\varepsilon})$ = $\{v^{d}_{ka}(I_a,\beta) - v^{e}_{ka}(I_a,\beta)\}/\mu_k(I_a,\beta)$, the scaled difference in the anticipated and realized net value for the chosen alternative. Even without the last assumption,

$$(51)\qquad \mathrm{UMCE^{e}}(s,\mathrm{k}) = \mathbf{E}_{\beta,\sigma|s,k}(v^{d}_{ka}(I_a,\beta) - v^{e}_{ka}(I_a,\beta))/\mu_k(I_a,\beta).$$

Selection again enters the calculation of UMCE for a class of consumers. For consumers with *s* ∈ **T** and observed "as is" choices in a set **D**,

$$(52)\qquad \mathrm{UMCE^{e}}(s,\mathrm{k}) = \frac{\mathbf{E}_{s|\mathbf{T}}\mathbf{E}_{\beta,\sigma|s}\sum_{k\in\mathbf{D}} L_{ka}(I_a,\beta,\sigma)\cdot[v^{d}_{ka}(I_a,\beta) - v^{e}_{ka}(I_a,\beta)]/\mu_k(I_a,\beta)}{\mathbf{E}_{s|\mathbf{T}}\mathbf{E}_{\beta,\sigma|s}L_{\mathbf{D}a}(I_a,\beta,\sigma)}.$$

These per capita transfers can be applied separately to disjoint **D** sets, or combined into a weighted average of the form (52) to give a uniform transfer for all consumers in **C** whose scenario *a* purchases are from **D**. Since only consumers who choose an alternative in subset **D** experience any difference between anticipated and realized net values, the numerator of (52) is the expected compensating variation per capita for all consumers with characteristics in **T**, while the denominator is the share of the population with characteristics in **T** and scenario *a* choices in **D**. In (56), commonly $v^{d}_{ka}(I_a,\beta,\sigma) \geq v^{e}_{ka}(I_a,\beta,\sigma)$ for all tastes. However, it is possible that there are tastes appearing in reality, or in the utility model approximation to it, that lead to some "as is" winners with $v^{d}_{ka}(I_a,\beta,\sigma) < v^{e}_{ka}(I_a,\beta,\sigma)$. This raises two issues, first whether the transfers should be calculated including or excluding winners in the calculation of the aggregate needed to make losers whole. The argument hinges on whether the distribution fulfilling the aggregate transfer can in principle claw back gains from winners to compensate losers; if not, the calculation should exclude winners. A related issue is that it may be impossible to distinguish winners and losers in the class of consumers in **C** and **D**, in which case the per capita calculation excluding winners but applied to both losers and winners gives an unwarranted transfer to winners.

In the second case, with false advertising or other misinformation about alternatives' actual attributes, the MCE is the difference between the realized utility obtained from (i) the alternative the person chose when misinformed and (ii) the alternative the person would have chosen if fully informed. If the chosen alternative is the same in the "but for" and "as is" scenarios, then $\mathrm{MCE}^{\mathrm{e}}(s,\mathrm{k},\beta,\sigma,\varepsilon) = 0$; i.e., there is no loss for consumers whose choice was unaffected by the misinformation. Since the "but for" anticipated net values are defined to match the net values that consumers *realized* in the "as is" situation, one has $\mathrm{v}^{\mathrm{e}}_{\mathrm{kb}}(I_{\mathrm{a}},\beta,\sigma) = \mathrm{v}^{\mathrm{d}}_{\mathrm{kb}}(I_{\mathrm{a}},\beta,\sigma) = \mathrm{v}^{\mathrm{e}}_{\mathrm{ka}}(I_{\mathrm{a}},\beta,\sigma)$ for all k. Given $\varepsilon_{\mathrm{a}} = \varepsilon_{\mathrm{b}}$, the experienced-utility MCE has the form (51) specialized to this relation among the net values:

$$\text{(53)}\quad \mathrm{UMCE}^{\mathrm{e}}(s,\mathrm{k})$$
$$= \mathbf{E}_{\beta,\sigma|s,\mathrm{k}}\left\{\max\left(\frac{\sigma}{L_{\mathrm{ja}}(I_{\mathrm{a}},\beta,\sigma)}\cdot\ln\frac{\sum_{\mathrm{k}=0}^{\mathrm{J_b}}\exp\left(\mathrm{v}^{\mathrm{e}}_{\mathrm{ka}}(I_{\mathrm{a}},\beta,\sigma)/\sigma\right)}{\sum_{\mathrm{k}=0}^{\mathrm{J_a}}\exp\left(\mathrm{v}^{\mathrm{d}}_{\mathrm{ka}}(I_{\mathrm{a}},\beta,\sigma)/\sigma\right)},0\right) + \mathrm{v}^{\mathrm{d}}_{\mathrm{ja}}(I_{\mathrm{a}},\beta,\sigma) - \mathrm{v}^{\mathrm{e}}_{\mathrm{ja}}(I_{\mathrm{a}},\beta,\sigma)\right\}/\mu(I_{\mathrm{a}},\beta,\sigma).$$

Again, it is normal in false advertising situations (but not necessarily for all forms of misinformation) that $\mathrm{v}^{\mathrm{e}}_{\mathrm{ka}}(I_{\mathrm{a}},\beta,\sigma) \le \mathrm{v}^{\mathrm{d}}_{\mathrm{ka}}(I_{\mathrm{a}},\beta,\sigma)$. Then (53) is less than (51); i.e., the transfer is lower when the "but for" scenario consists of providing the correct information that leads anticipated and realized utilities to agree than when the "but for" scenario consists of providing consumers with their anticipated utilities. When there are tastes such that $\mathrm{v}^{\mathrm{e}}_{\mathrm{ka}}(I_{\mathrm{a}}) > \mathrm{v}^{\mathrm{d}}_{\mathrm{ka}}(I_{\mathrm{a}})$, so that these consumers win from the misrepresentation, there is again a question of whether they should be included or excluded in the calculation of the per capita transfer.

Analogously to (52), in the class of consumers with characteristics in **T** who chose alternative J in scenario *a*,

$$\text{(54)}\quad \mathbf{E}_{\varepsilon|\beta,\sigma,,\delta_{\mathrm{Ja}}(I_{\mathrm{a}})=1}\mathrm{MCE}^{\mathrm{e}}(\beta,\sigma,\varepsilon)$$
$$= \frac{\mathbf{E}_{s|\mathbf{T}}\mathbf{E}_{\zeta|s}\mathrm{P}_{\mathrm{Ja}}(I_{\mathrm{a}},\beta,\sigma)\cdot\left(\mathrm{v}^{\mathrm{a}}_{\mathrm{j1}}(I_{\mathrm{a}},\beta) - \mathrm{v}^{\mathrm{r}}_{\mathrm{j1}}(I_{\mathrm{a}},\beta)\right) + \mathbf{E}_{\varepsilon|\beta,\sigma,,\delta_{\mathrm{Ja}}(I_{\mathrm{a}})=1}\max\left(\sigma\cdot\ln\frac{\sum_{\mathrm{k}=0}^{\mathrm{J_2}}\exp\left(\mathrm{v}^{\mathrm{e}}_{\mathrm{ka}}(I_{\mathrm{a}},\beta)/\sigma\right)}{\sum_{\mathrm{k}=0}^{\mathrm{J_1}}\exp\left(\mathrm{v}^{\mathrm{d}}_{\mathrm{ka}}(I_{\mathrm{a}},\beta)/\sigma\right)},0\right)/\mu(I_{\mathrm{a}},\beta,\sigma)}{\mathbf{E}_{\varepsilon|\beta,\sigma,,\delta_{\mathrm{Ja}}(I_{\mathrm{a}})=1}\mathrm{P}_{\mathbf{Ja}}(I_{\mathrm{a}},\beta,\sigma)}.$$

Retrospective welfare analysis for consumer durables whose attributes are affected by contract violations or deceptions can require a combination of the preceding calculations. For example, consider homeowners whose properties lose value due to groundwater contamination from an industrial site, or automobile owners whose vehicles fail to deliver promised performance after correction of defective emission controls, and lose resale value as a result. Then members of the class of owners of the affected durables at the time the defect is announced are harmed in the amount given by (51) if they are legally entitled to a non-defective product, as in the case of environmental injury, or given by (53) if they are legally entitled only to the opportunity to make a product choice with the correct information, as in the case of false advertising. Further, as long as there is no further contract violation or deception following the announcement, the harm is fully capitalized in the resale value of the durables and these calculations conclude the calculation of harm. Pre-announcement owners who choose to continue to hold their durables have willingly declined the opportunity to mitigate their losses by selling, and post-announcement buyers who find that the lower price offsets the reduced performance are not harmed.

8. AN ILLUSTRATIVE APPLICATION

An empirical example of applied welfare analysis using the methods of this paper, due to Kenneth Train (2015), examines the impact on consumers of video streaming services that share customers' personal and usage information without their prior knowledge. This analysis is based on choice models estimated using data from a conjoint experiment designed and described by Butler and Glasgow (2015). Each choice experiment included four alternative video steaming services with specified price and the attributes listed in Table 4 plus a fifth alternative of not subscribing to any video streaming service.

Each of 260 respondents was presented with 11 choice experiments. The choice model was of the form (9) for money-metric utility, with (β,ln σ) having a multivariate normal distribution. Estimates obtained using maximum simulated likelihood are given in Table 5. The results indicate that people are willing to pay \$1.56 per month on average to avoid commercials. Fast availability is valued highly, with an average WTP of \$3.95 per month in order to see TV shows and movies soon after their original showing. On average, people prefer having a mix with more TV shows and fewer movies, but the mean is not significantly different from zero. Average willingness to pay for more content of both kinds is \$2.96 per month. Interestingly, people who want fast availability tend to be those who prefer more TV shows and fewer movies: the correlation between these two WTP's is 0.51, while the correlation between WTP for fast availability and more content of both kinds is only 0.04. Apparently, the desire for fast availability mainly applies to TV shows.[18]

[18] The model was also estimated using an Allenby-Train hierarchical Bayes method, with similar results; the details of both estimation methods are given in Bhat (2001); Train (2000, 2009, 2015), and Ben Akiva, McFadden, and Train (2016).

Table 4. Non-Price Attributes

Attribute	Levels
Commercials shown between content	Yes ("commercials') No (baseline category)
Speed of content availability	TV episodes next day, movies in 3 months ("fast content") TV episodes in 3 months, movies in 6 months (baseline category)
Catalogue	10,000 movies and 5,000 TV episodes ("more content") 2,000 movies and 13,000 TV episodes ("more TV/fewer movies") 5,000 movies and 2,500 TV episodes (the baseline category)
Data-sharing policies	Information is collected but not shared (baseline category) Usage information is share with third parties ("share usage")[19] Usage and personal information are shared with third parties ("share usage and personal")

Table 5A. MSL Estimates of WTPs for Video Streaming Services

	Population Mean		Std Dev in Population	
	Estimate	Std Error	Estimate	Std Error
Ln(1/σ)	-2.002	0.0.945	1.0637	0.0755
WTP for:				
Commercials	-1.562	0.4214	3.940	0.5302
Fast Availability	3.945	0.4767	3.631	0.4138
More TV, fewer movies	-0.6988	0.4783	4.857	0.5541
More content	2.963	0.4708	2.524	0.4434
Share usage only	-0.6224	0.4040	2.494	0.4164
Share personal and usage	-2.705	0.5844	6.751	0.7166
No service	-27.26	2.662	19.42	2.333

Table 5B. Correlation Point Estimates

(* denotes significance at 5% level)

	Commer-cials	Fast Avail-ability	Mostly TV	Mostly movies	Share usage	Share personal and usage	No service
Ln(1/σ)	-0.5813*	-0.1371	0.0358	0.0256	0.0022	-0.1287	0.2801*
Commercials	1.0000	0.1172	-0.3473*	0.0109	-0.2562	-0.0079	-0.4108*
Fast Availability		1.0000	0.8042*	-0.4019*	-0.3542*	-0.4206*	0.2391*
Mostly TV			1.0000	-0.5890*	-0.1695	-0.3328*	0.4616*
Mostly movies				1.0000	0.5141*	0.5181*	-0.0147
Share usage					1.0000	0.9370*	-0.0563
Share personal and usage						1.0000	-0.0975
No service							1.0000

[19] Butler and Glasgow use the terms "non-personally identifiable information (NPPI)" and "personally identifiable information (PII)" for what we are labelling "share usage" and "share usage and personal".

Consider how a video streaming service might share its subscribers' personal and usage information with third parties who then use that information for targeted marketing to the subscribers. The Table 5 estimates imply that consumers have an average WTP of 62 cents per month to avoid having their usage data shared in aggregate form; however, the hypothesis of zero average WTP cannot be rejected. Consumers are much more concerned about their personal information being shared along with their usage information: The average WTP to avoid such sharing is $2.71 per month. The correlation between WTP to avoid the two forms of sharing is a substantial 0.937. However, some people like having their data shared, because they value the targeted marketing that they receive as a result of the sharing. In the demand model, the WTP is normally distributed with a mean of -2.71 and standard deviation of 6.751, which implies that 34.4% of the population like to have their information shared.

For the welfare analysis, there are three providers, Netflix, Amazon Prime, and Hulu, and that customers can subscribe to any one of these services, any combination of them, or to no service. Table 6 gives the "as is" alternatives available to customers, and the shares of customers in the sample who chose each alternative. At the time of the survey, Hulu had about 6 million subscribers, which, given the market shares above, imply that total market size is 31 million potential subscribers. This is less than the number of households in the US because the survey screened for people who either already subscribe, or were likely to subscribe, to a video-screening service if they did not currently have one. The market is then the US households who are open to the possibility of subscribing to a video streaming service.

Table 6: Market Shares of Video Steaming Service Portfolios

Alternative	Share
Netflix	0.2867
Amazon Prime	0.0467
Hulu	0.0400
Netflix + Amazon Prime	0.1167
Netflix + Hulu	0.0700
Amazon Prime + Hulu	0.0100
Netflix + Amazon Prime + Hulu	0.0733
No video streaming service	0.3567

In the "as is" scenario, customers think that none of the service providers shares their usage and personal information, but in fact one of them does. The analysis chooses Hulu as the one who shares, but the selection is arbitrary. How much are consumers hurt by the fact that Hulu shared its subscribers information without their knowing beforehand, and how much would Hulu be liable for under different theories of damages?

Assume for the welfare analysis that when people were choosing among services, they anticipated that these services would have the attributes given in Table 7. Note that none of the providers were thought to share their subscribers' information.

Table 7: Anticipated Attributes for Decision Utility

	Netflix	Amazon Prime	Hulu
Price per month	7.99	6.58	7.99
Commercials	0	0	0
Fast Availability	0	0	1
More TV, fewer movies	0	1	0
More content	1	0	0
Share usage only	0	0	0
Share personal and usage	0	0	0

The attributes of the alternatives that represent multiple services are the sum of the attributes of the services within the packages. For example, the price of Netflix+Amazon Prime is $14.67 per month and provides the "More content" of Netflix and the "MoreTV, fewer movies" of Amazon Prime. Alternative specific constants were calibrated such that the predicted shares for the alternatives equal the observed shares in Table 7.

Now suppose that, in reality, Hulu shared its subscribers' personal and usage information, and that this fact was revealed months after people began subscribing. The experienced utility is based on the attributes in Table 7 except that "Share personal and usage" receives a 1 for Hulu. What is the difference between the welfare that people expected to obtain when they made their choices compared to the welfare they actually obtained? Only Hulu subscribers obtained experienced utility that differed from decision utility. The aggregate difference is $22.9 million per month, or $3.81 on average for Hulu subscribers. Note that, for the population as a whole, the average WTP to avoid sharing is $2.71, as stated above. The average WTP conditional on having subscribed to Hulu is $3.81. That is, the average Hulu subscriber dislikes sharing their information more than the average person in the population does. How does this arise? Note in Table 5B that the correlation between the WTPs for between "Fast Availability" and Share personal and usage" is -0.42. Hulu is the only service that offered Fast Availability, and so people who valued this attributed tended to choose Hulu. However, the people who place a high value on Fast Availablity also tend to dislike sharing their information more than other people. The difference between the conditional mean of $3.80 and the unconditional mean of $2.71 arises because of this correlation.

The damages that Hulu would need to pay in compensation for its sharing of its subscribers' information depends critically on what was illegal: was it illegal for Hulu to share its customers' information, or was it illegal for Hulu not to disclose that it was doing so. If it was illegal for Hulu to share its subscribers' information, then the aggregate damage that Hulu is responsible for is $22.9 million for each month that the sharing had been undisclosed. However, some customers like having their data shared, and this aggregate nets their gains from the losses that people who dislike sharing incurred. To obtain Pareto neutral compensation on a person-by-person basis, the $22.80 would not be enough to compensate the people who were hurt by the sharing: the people who liked the sharing would need to contribute their gains too. We can calculate the welfare impact separately for people who like sharing and people who dislike sharing. Among the Hulu subscribers who have a negative WTP for sharing, the aggregate loss in welfare is $30.4 million. Hulu subscribers who have a positive WTP for sharing obtained an aggregate gain of $7.50 million. For Hulu to be able to compensate the people who were hurt from its sharing, Hulu would need to pay $30.4, since it does not have the ability to claw back compensation from the people who gained.

Next suppose information sharing is legal, but nondisclosure is Illegal. If Hulu is liable for nondisclosure, then the relevant comparison is between

(i) the utility that consumers obtained in the "as is" situation, where they choose among the alternative under the concept that Hulu did not share but it in fact did; this is the realized utility for the alternative that the person chose based on decision utilities, and

(ii) the utility that consumers would have obtained Hulu had disclosed its sharing practice before customers choose among the services; this is the realized utility that the customer would choose based on realized utilities.

Every Hulu subscriber who likes sharing would have chosen Hulu if they had known in advance that it shared information. And some of the Hulu subscribers who dislike sharing would still have chosen Hulu if they had known that Hulu shared their information. None of these subscribers were hurt by the nondisclosure. The only Hulu subscribers who were hurt by the nondisclosure are those who dislike sharing sufficiently that they would not have chosen Hulu if they had known the sharing practice. However, the welfare losses from non-disclosure are not borne only by Hulu subscribers. People who like sharing but didn't know that Hulu shares and chose a different provider were potentially hurt because they were not able to take advantage of this undisclosed attribute of Hulu service. People who would have chosen Hulu if they had known that Hulu shares but didn't obtained less welfare than they would have obtained under full disclosure. Table 8 gives the losses for each group of consumers from the non-disclosure of Hulu's sharing practice.

Table 8: Damages Arising from Non-Disclosure

	Aggregate loss, in million $ per month	Average loss per person in the market	Average loss for Hulu subscribers	Average loss for people who did not subscribe to Hulu
All people	16.5	0.53	2.16	0.14
People who dislike sharing	13.0	0.64	3.05	0.00
People who like sharing	3.5	0.33	0.00	0.39

The total loss is $16.5 million per month, which consists of $13.0 million loss to people who dislike sharing and 3.55 loss to people who like sharing. The $13.0 million loss was incurred by Hulu subscribers who dislike sharing sufficiently to not choose Hulu if they had known its sharing practices. The $3.5 million loss was incurred by people who did <u>not</u> subscribe to Hulu but like sharing sufficiently to have chosen Hulu if they had known its sharing practices. The average loss per person in the population is simply the aggregate loss divided by market size (31 million). The average loss for Hulu subscribers can best be explained by starting in the bottom row of Table 10. Hulu subscribers who like sharing their information incurred zero harm from the nondisclosure: they subscribed to Hulu and so obtained the benefits of the sharing even though they didn't realize beforehand that they would. Importantly, they also did not gain from the nondisclosure. They obtained greater welfare from Hulu than they had expected when they chose Hulu. But they obtained the benefits of sharing even without prior disclosure, which would not have changed anything for them. Hulu subscribers who dislike sharing were hurt by $3.05 on average. Not all Hulu subscribers who dislike sharing were hurt by the non-disclosure. Only those who would not have chosen Hulu if they had known of its sharing practices were hurt, and these people were hurt by more than $3.05 on average (since the $3.05 average include Hulu subscribers who dislike sharing but were not hurt from the nondisclosure since they still would have chosen Hulu.) The top row in Table 10 gives a loss per Hulu subscriber of $2.16: it is the average of the $3.05 in the second row and $0.00 in the third row, weighted by the share of Hulu subscribers who dislike and like sharing. The losses for people who did not subscribe the Hulu are analogous. People who dislike sharing and did not subscribe to Hulu incurred no loss, since they would not have chosen Hulu if its sharing practices had been disclosed. Some people who did not subscribe to Hulu but like sharing would have chosen Hulu if they had known that Hulu shared their information. These people obtained less utility that they could have obtained under full disclosure.

In the "as is" situation, 19.3 percent of people in the market subscribed to Hulu. If everyone had been informed about Hulu's sharing practice, then this share would have dropped to 16.0 percent, which is a 17 percent reduction in subscribers. However, as explained above, this change includes two different movements: the share drops because some Hulu subscribers would not have chosen Hulu if they had known that Hulu would share their information, and the share rises because some people who did not subscribe to Hulu would have subscribed if they had known. Table 9 gives the share of people in each group. 12.5% of people subscribed to Hulu and would still have also done so if the sharing practice had been disclosed. 6.8% subscribed to Hulu but would not have if they had known about its sharing practice. That is, about a third of Hulu's subscribers would have not subscribed if they had been informed. 3.5% of people did not subscribe to Hulu but would have done so if they had known that Hulu shares their information.

Table 9: Choice Shares without and with Disclosure

	Would have subscribed to Hulu if its sharing practices had been disclosed	Would not have subscribed to Hulu if its sharing practices had been disclosed	Total
Subscribed to Hulu	0.125	0.068	0.193
Did not subscribe to Hulu	0.035	0.772	0.807
Total	0.160	0.840	

The share of people who subscribed to Hulu was 19.3%. If its sharing practices had been disclosed, then the share of subscribers would have been 0.193-0.068+0.035 = 0.16, i.e., 16 % as stated above.

9. CONCLUSIONS

This paper provides a foundation for applied welfare analysis of product regulation or compensation for product defects. It gives a practical setup for money-metric indirect utility functions whose features can be estimated using data on choice in real or hypothetical markets, and shows that there is essentially no loss of generality in restricting analysis to this setup. It draws a distinction between prospective and retrospective policy applications, and between cases where compensating transfers are hypothetical or are actually fulfilled. It introduces a Market Compensating Equivalent (MCE) welfare measure, an updated version of Marshallian consumer surplus, and shows that when compensating transfers are not actually fulfilled, it is preferred to commonly prescribed Hicksian compensating or equivalent variations. Further, MCE is shown to have desirable computational and aggregation properties. The problem of carrying out welfare calculations when tastes of individual consumers are only partially observed is addressed, and computational formulas are given for calculation of expected compensating transfers. Decision-utility and experienced-utility are distinguished, and the issues of conducting welfare calculus in experienced utility are discussed. A number of common welfare calculus problems are treated, and formulas are given for their resolution. Finally, an application illustrates the use of these methods and the importance of the distinctions introduced in this paper.

REFERENCES

Afriat, S. (1967) “The construction of utility functions from expenditure data,” *International Economic Review*, 8, 67-77.

Alexandrov, A. (1939) “Almost everywhere existence of the second differential of a convex function and surfaces connected with it,” *Lenningrad State University Annals*, Mathematics Series 6;3-35.

Aliprantis, C.; K. Border (2006) *Infinite dimensional Analysis*, Springer: Berlin.

Allcott, H., (2013) “The welfare effects of misperceived product costs: data and calibrations from the automobile market.” *Am. Econ. J.: Econ. Policy* 5 (3), 30–66.

Anas, A. and C. Feng (1988) “Invariance of Expected Utilities in Logit Models,” *Economic Letters* 27:1, 41-45.

Arrow, K. (1950) “A difficulty in the concept of social welfare,” *Journal of Political Economy*, 58.4, 328-346.

Ben-Akiva, M.; D. McFadden: K. Train (2016) “Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-Based Conjoint Analysis,” University of California, Berkeley, working paper.

Bentham, J. (1789) *An introduction to the principles of morals and legislation*, Oxford: The Clarendon Press, 1876.

Bergson, A. (1938) “A reformulation of certain aspects of welfare economics,” *Quarterly Journal of Economics*, 52.2, 310-334.

Bernheim, D. (2016) “The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics,” *Journal of Benefit Cost Analysis*,, 7.1, 12-68.

Bhattacharya, D. (2015) “Nonparametric Welfare Analysis for Discrete Choice,” *Econometrica*, 83.2, 617-649.

Bhattacharya, D. (2017) “Empirical Welfare Analysis for Discrete Choice: Some General Results,” Cambridge University working paper.

Blackorby, C.; R. Boyce; R. Russell (1978) “Estimation of demand systems generated by the Gorman Polar Form,” *Econometrica*, 46, 345-364.

Border, K. (2014) “Monetary Welfare Measurement,” Cal Tech lecture notes.

Chipman, J.; J. Moore (1980) "Compensating Variation, Consumer's Surplus, and Welfare," American Economic Review. 70: 933-49

Chipman, J.; J. Moore (1990) "Acceptable Indicators of Welfare Change, Consumer's Surplus Analysis, and the Gorman Polar Form," in D. McFadden, M. Richter (eds) *Preferences, uncertainty, and optimality: Essays in honor of Leonid Hurwicz*. Boulder and Oxford: Westview Press; 68-120.

Chorus, C.G., H. Timmermans (2009) “Measuring user benefits of changes in the transport system when traveler awareness is limited,” *Transportation Research Part A ,* 43(5), 536-547.

Conniffe, D. (2007) “A Note on Generating Globally Regular Indirect Utility Functions,” *Journal of Theoretical Economics*, 7.1, 1-11.

Dagsvik, J.; 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.

Deaton, A.; J. Muellbauer (1980) “An almost ideal demand system,” *American Economic Review*, 70, 312-326.

Debreu, G. (1959) *Theory of Value*, New Haven : Yale University Press.

Diamond, P. and D. McFadden (1974), “Some uses of the expenditure function in public finance,” *Journal of Public Economics* 3.1 3-21.

Doha, E. H.; A. H. Bhrawy; M. A. Saker (2011) “On the Derivatives of Bernstein Polynomials,” *Boundary Value Problems*, doi:10.1155/2011/829543 p. 1-16.

Dubin, J. (1985) *Consumer Durable Choice and the Demand for Electricity*, Elsivier: New York.

Dubin, J., D. McFadden (1984) "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345-62

Dudley, R. (2002) *Real Analysis and Probability*, Cambridge University Press, New York.

Dunford, N.; J. Schwartz (1964) *Linear Operators*, Interscience, New York.

Dupuit, J. (1844) "On the Measurement of the Utility of Public Works", *Annales des ponts et chaussées*. (English translation, *International Economic Review*, 1952).

Edgeworth, F. Y. (1881) *Mathematical Psychics; an essay on the application of mathematics to the moral sciences*,, London, C. K. Paul & Co.
Fosgerau, M.; D. McFadden; M. Bierlaire (2013) “Choice Probability Generating Functions,” *Journal of Choice Modelling*, 8, 1-18.
Fosgerau, M.; D. McFadden (2012) “A theory of the perturbed consumer with general budgets,” working paper.
Gorman, W. (1953) “Community Preference Fields,” *Econometrica*, 21, 63-80.
Gorman, W. (1961) “On a Class of Preference Fields,” *Metroeconomica*, 13, 53-56.
Gossen, H. (1854) *Die Entwicktlung*, English translation: *The Laws of Human Relations*, Cambridge: MIT Press, 1983.
Hall, P.; A. Yatchew (2007) “Nonparametric Estimation when Data on Derivatives are Available,” *The Annals of Statistics*, 35.1, 300-323.
Hammond, P. (1994) "Money Metric Measures of Individual and Social Welfare Allowing for Environmental Externalities," in W. Eichhorn (ed) *Models and Measurement of Welfare and Inequality*, Springer-Verlag, 694-724.
Heiss, F.; D. McFadden; J. Winter (2013) “Plan Selection in Medicare Part D: Evidence from Administrative Data,” *Journal of Health Economics* 32.6, 1325-1344.
Hicks, J. (1939) *Value and Capital*, Oxford, Clarendon press.
Houthakker, H. (1950) “Revealed preference and the utility function,” *Economica*, N.S. 17, 159-174.
Hurwicz, L.; H. Uzawa (1971) "On the Integrability of Demand Functions," in J. Chipman, L. Hurwicz, M. Richter, and H. Sonnenschein (eds) *Preferences, Utility, and Demand*, New York: Harcourt, 114-148.
Jevons, W. (1871) *Theory of Political Economy*, reprinted by London, Macmillan, 1931.
Jorgenson, D. (1997) *Welfare*, MIT Press: Cambridge, Vol. 1 and 2.
Johnson, N. and S. Kotz (1970, Ch. 21) *Continuous Univariate Distributions-1*, Houghton-Mifflin: New York.
Kadison, R. and Z. Liu (2016) *Bernstein Polynomial and Approximation*, lecture notes.
Kaldor, N. (1939) "Welfare Propositions of Economics and Interpersonal Comparisons of Utility," Econ. Jour., XLIX, 549-52.
Katzner, Donald (1970) *Static Demand Theory*, New York: Macmillan.
Kosorok, M. (2008) *Introduction to Empirical Processes and Semiparametric Inference*, Springer: New York.
Lorentz, G. (1937) “Zur theorie der polynome von S. Bernstein,” Matematiceskij Sbornik 2, 543–556.
Lowenstein, G.; Ubel (2008) “Hedonic adaptation and the role of decision and experience utility in public policy,” *Journal of Public Economics*, 92, 1795-1810.
Marshall, A. (1890) *Principles of Economics*, London: Macmillan.
Mas-Colell, A.; M. Whinston, and J. Green (1995) *Microeconomic Theory*, Oxford: Oxford University Press.
Matzkin, R. and D. McFadden (2011) “Trembling Payoff Market Games,” working paper.
McFadden, D. (1974) “The Measurement of Urban Travel Demand,” *Journal of Public Economics*, 3, 303-328.
McFadden, D. (1981) “Structural Discrete Probability Models Derived from Theories of Choice,” in C. Manski and D. McFadden (eds) *Structural Analysis of Discrete Data and Econometric Applications*, MIT Press: Cambridge, 198-272.
McFadden, D. (1986) “The Choice Theory Approach to Market Research,” *Marketing Science*, 275-297.
McFadden, D. (1994) "Contingent valuation and social choice," *American Journal of Agricultural Economics* 76, 689-708.
McFadden, D. (1999) "Computing Willingness-to-Pay in Random Utility Models," in J. Moore, R. Riezman, and J. Melvin (eds.), *Trade, Theory, and Econometrics: Essays in Honour of John S. Chipman*, Routledge: London.
McFadden, D. (2004) “Welfare Economics at the Extensive Margin: Giving Gorman Polar Consumers Some Latitude,” University of California, Berkeley, working paper.
McFadden, D. (2008) “Environmental Valuation of Environmental Projects,” Univ. of California working paper.
McFadden, D. (2012) “Economic Juries and Public Project Provision,” *Journal of Econometrics*, 166, 116-126.
McFadden, D. (2014) “The New Science of Pleasure: Consumer Behavior and the Measurement of Well-Being,” in S. Hess and A. Daly, eds, *Handbook of Choice Modelling*, Elgar: Cheltenham, 7-48.
McFadden, D. (2017) “Stated Preference Methods and their Applicability to Environmental Use and Non-Use Valuations,” in D. McFadden and K. Train (eds) *Contingent Valuation of Environmental Goods: A Comprehensive Critique*, Elgar: Cheltingham, Chap. 6.
McFadden, D., K. Train (2000) "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447-470.
McFadden, D.; B. Zhou (2015) “Measuring Lost Welfare from Poor Health Insurance Choices,” Schaeffer Center, USC.
Miller, K.; et al. (2011) “How Should Consumers’ Willingness to Pay Be Measured? An Empirical Comparison of State-of-the-Art Approaches,” *Journal of Marketing Research*, 48.1, 172-184.
Pareto, Vilfredo (1906) *Manual of Political Economy*, English Translation, Augustus M. Kelley, NY, 1971.

Peleg, B. (1970) “Utility functions for partially ordered topological spaces,” *Econometrica*, 38, 93-96.
Pollard, D. (1984) *Convergence of Stochastic Processes*, Springer, New York.
Rademacher, H. (1919) “Ubër partielle und totale Differenzierbarkeit von Funktionen mehrerer Variabeln und ubër die Transformation der Doppelintegrale,” *Math. Ann.* 79, 340-359.
Rader, T. (1973) “Nice demand functions,” *Econometrica*, 41, 913-935.
Resnic, S.; R. Roy (1990) “Leader and Maximum Independence for a Class of Discrete Choice Models,” *Economic Letters,* 33.3, 259-263.
Richter, M. (1966) “Revealed Preference Theory,” *Econometrica*, 34, 635-645.
Roy, R. (1947) "La Distribution du Revenu Entre Les Divers Biens". *Econometrica*, 15.3, 205–225.
Samuelson, P. (1947) *Foundations of economic analysis*, Cambridge: Harvard University Press, 1983.
Samuelson, P. (1948) “Consumption theory in terms of revealed preference,” *Economica*, 15, 243-253.
Schmeiser, S. (2014) “Consumer inference and the regulation of consumer information,” *Int. J. Ind. Organ.* 37, 192–200.
Scitovsky, T. (1951) “The State of Welfare Economics,” *American Economic Review*, 41-3, 303-315.
Sen, A. (2017) *Collective Choice and Social Welfare*, Harvard University Press: Cambridge.
Shannon, C. (2006) “A Prevalent Transversality Theorem for Lipschitz Functions,” *Proceedings of the American Mathematical Society*, 134.9, 2755-2755.
Slutsky, E. (1915) "Sulla teoria del bilancio del consummatore", *Giornale degli Economisti*. English translation, "On the Theory of the Budget of the Consumer," in G. Stigler and K. Boulding, eds, *Readings in Price Theory*, Homewood: Irving.
Small, K.; S. Rosen (1981), “Applied Welfare Economics with Discrete Choice Models,” *Econometrica*, 49.1, 105-130.
Smith, A. (1776) *An inquiry into the nature and causes of the wealth of nations*. London, W. Strahan and T. Cadell.
Thaler, R.; C. Sunstein (2003) “Libritarian Paternalism,” *American Economic Review*, 93.2, 175-179.
Thaler, R.; C. Sunstein (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Yale University Press: New Haven.
Thurstone, L. (1927) “A Law of Comparative Judgment,” *Psychological Review*, 34: 273-286.
Train, K. (2015) “Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples,” *Journal of Choice Modelling*, 16, 15-22.
Van der Vaart, A.; J. Wellner (1996) *Weak Convergence and Empirical Processes*, Springer, New York.
Varian, H. (1982) "The Nonparametric Approach to Demand Analysis," *Econometrica,* 50: 945-73.
Varian, H. (2006) *Revealed Preference*. New York: Oxford University Press.
Willig, Robert (1976) "Consumer's Surplus without Apology," *American-Economic-Review*. 66: 589-97.
Yatchew, A. (1985) “Applied Welfare Analysis with Discrete Choice Models: Comment,” *Economic Letters*, 18.1, 13-16.
Zhao, Y.; K. Kockelman; A. Karlstrom (2012) “Welfare Calculations in Discrete Choice Settings: The Role of Error Term Correlation,” *Transport Policy* 19.1, 76-84.

Appendix A: Approximation Theory for Functions and Probabilities

This appendix provides the mathematical basis for uniform parametric approximations to utility functions and probabilities. The first theorem adapts Bernstein-Weierstrauss approximation theory to the class of functions considered in this paper, and the second theorem utilizes Pollard's methods for establishing uniform weak convergence of empirical probabilities; see Lorentz (1937), Kadison-Liu (2016), Pollard (1984).

Let $b_{jK}(p) = \binom{K}{j} p^j(1-p)^{K-j}$ denote the binomial probability of j successes in K draws, each with probability $p \in [0,1]$; and define $b_{j,K}(p) \equiv 0$ for $j < 0$ or $j > K$. Differentiating, $\frac{d}{dp} b_{jK}(p) = K[b_{j-1,K-1}(p) - b_{j,K-1}(p)]$. Higher order derivatives can be defined recursively; see Doha et al (2011). Note that $\sum_{j=0}^{K} b_{jK}(p) \equiv 1$ and $\sum_{j=0}^{K} \frac{d}{dp} b_{jK}(p) \equiv 0$. The following result is a straightforward multivariate restatement of the Bernstein-Weierstrauss theorem on approximation of continuous functions by polynomials.

Theorem A.1. Let H denote a compact metric space with metric h. Consider $f \in \mathfrak{C}([0,1]^n \times H)$. Let $\mathbf{K} = (K_1,\ldots,K_n)$ denote a vector of positive integers, $\mathbf{j} = (j_1,\ldots,j_n)$ a vector of integers satisfying $0 \le j_i \le K_i$ for $i = 1,\ldots,n$, $\mathbf{p} = (p_1,\ldots,p_n) \in [0,1]^n$, $\mathbf{j}\oslash\mathbf{K} = (j_1/K_1,\ldots,j_n/K_n)$. Define the multivariate binomial probability $b_{\mathbf{j},\mathbf{K}}(\mathbf{p}) = \prod_{i=1}^{n} b_{j_iK_i}(p_i)$, the vector $\beta_{\cdot\mathbf{K}}(h)$ of functions $\beta_{\mathbf{j},\mathbf{K}}(h) \equiv f(\mathbf{j}\oslash\mathbf{K},h)$ on H for $\mathbf{0} \le \mathbf{j} \le \mathbf{K}$, and the multivariate polynomial $B_{\mathbf{K}}(\mathbf{p},\beta_{\cdot\mathbf{K}}(h)) = \sum_{\mathbf{0}\le\mathbf{j}\le\mathbf{K}} b_{\mathbf{j},\mathbf{K}}(\mathbf{p})\beta_{\mathbf{j},\mathbf{K}}(h)$. Let C denote the compact range of f. Then, $\beta_{\mathbf{j},\mathbf{K}} \in \mathfrak{C}(H,C))$, and $B_{\mathbf{K}}(\mathbf{p},\beta_{\cdot\mathbf{K}}(h))$ has the following approximation properties: (i) $\lim_{\mathbf{K}\to\infty} \max_{[0,1]^n\times H} |B_{\mathbf{K}}(\mathbf{p},\beta_{\cdot\mathbf{K}}(h)) - f(\mathbf{p},h)| = 0$, and (ii) if $\partial f(\mathbf{p},h)/\partial p_i$ exists and is continuous on a closed set $A \subseteq [0,1]^n \times H$, then $\lim_{\mathbf{K}\to\infty} \max_{A} \left|\frac{\partial B_{\mathbf{K}}(\mathbf{p},\beta_{\mathbf{K}}(h))}{\partial p_i} - \frac{\partial f(\mathbf{p},h)}{\partial p_i}\right| = 0$. If in addition, f is Lipschitz in its arguments, then $\beta_{\cdot\mathbf{K}}$ is Lipschitz on H.

Proof: The continuous function f is uniformly continuous on $[0,1]^n \times H$ and bounded by a constant M, so that given $\varepsilon > 0$, there exists $\delta \in (0,1)$ such that $|\mathbf{p}' - \mathbf{p}| \le \delta$ and $h(h,h') \le \delta$ imply $|f(\mathbf{p}',h') - f(\mathbf{p},h)| < \varepsilon/6$. Define the set $\mathbf{J}_\delta = \{\mathbf{j} | \mathbf{0} \le \mathbf{j} \le \mathbf{K} \text{ and } |\mathbf{j}\oslash\mathbf{K} - \mathbf{p}| \le \delta/2\}$. By Hoeffding's inequality, $\text{Prob}(\mathbf{J}_\delta^c) \le 2\sum_{i=1}^{n} \exp(-\delta^2 K_i/2)$. Select $\mathbf{K} \ge 192nM/\varepsilon\delta^3$ and $\mathbf{K} > 2/\delta$. In the inequality $|B_{\mathbf{K}}(\mathbf{p},\beta_{\cdot\mathbf{K}}(h)) - f(\mathbf{p},h)| \le \left\{\sum_{\mathbf{J}_\delta} + \sum_{\mathbf{J}_\delta^c}\right\} b_{\mathbf{j},\mathbf{K}}(\mathbf{p})|f(\mathbf{j}\oslash\mathbf{K},h) - f(\mathbf{p},h)|$, the first sum is bounded by $\varepsilon/6$, while the second sum is bounded by $2M\cdot\text{Prob}(\mathbf{J}_\delta^c) \le 4M\cdot\sum_{i=1}^{n} \exp(-\delta^2 K_i/2) \le 4M\cdot\sum_{i=1}^{n} \frac{2}{\delta^2 K_i} \le \varepsilon\delta/24 \le \varepsilon/24$. This establishes $|B_{\mathbf{K}}(\mathbf{p},\beta_{\cdot\mathbf{K}}(h)) - f(\mathbf{p},h)| < \varepsilon/3$ for each $(\mathbf{p},h) \in [0,1]^n \times H$.

Next suppose that on a compact set A, $\partial f(\mathbf{p},h)/\partial p_1$ exists and is continuous. Then it is uniformly continuous and bounded on A; let M be a bound. The δ above can be chosen so that $\left|\frac{\partial f(\mathbf{p}',h')}{\partial p_1} - \frac{\partial f(\mathbf{p},h)}{\partial p_1}\right| \le \frac{\varepsilon}{6}$ and $\frac{f(\mathbf{p}+\delta'\Delta_1,h) - f(\mathbf{p},h)}{\delta'} - \frac{\partial f(\mathbf{p},h)}{\partial p_1} \equiv \zeta(\delta',\mathbf{p},h)$ with $|\zeta(\delta',\mathbf{p},h)| \le \frac{\varepsilon}{6}$ for $|\delta'| \le \delta$ and $|\zeta(\delta',\mathbf{p},h)| \le M(1+2/\delta)$ for $|\delta'| > \delta$, where Δ_1 is a vector with a one in component 1, zeros elsewhere. Define $\mathbf{p}_{2+} = (p_2,\ldots,p_n)$, $\mathbf{j}_{2+} = (j_2,\ldots,j_n)$, $\mathbf{K}_{2+} = (K_2,\ldots,K_n)$, and $b_{\mathbf{j}_{2+},\mathbf{K}_{2+}}(\mathbf{p}_{2+}) = \prod_{i=2}^{n} b_{j_iK_i}(p_i)$ on $[0,1]^{n-1}$. Then

$$\frac{\partial B_K(\mathbf{p},\beta_K(h))}{\partial p_i} = K_1\left\{\sum_{J_\delta} + \sum_{J_\delta^c}\right\} b_{\mathbf{j}_{2+},\mathbf{K}_{2+}}(\mathbf{p}_{2+})\left(b_{\mathbf{j_1-1},\mathbf{K_1-1}}(p_1) - b_{\mathbf{j_1},\mathbf{K_1-1}}(p_1)\right) f(j_1/K_1\,,\mathbf{j}_{2+}\oslash\mathbf{K}_{2+},h)$$

$$= \left\{\sum_{\mathbf{J}_\delta} + \sum_{\mathbf{J}_\delta^c}\right\} b_{\mathbf{j}_{2+},\mathbf{K}_{2+}}(\mathbf{p}_{2+}) b_{\mathbf{j_1},\mathbf{K_1-1}}(p_1)\left[\frac{f((j_1+1)/K_1,\mathbf{j}_{2+}\oslash\mathbf{K}_{2+},h) - f(j_1/K_1,\mathbf{j}_{2+}\oslash\mathbf{K}_{2+},h)}{1/K_1}\right]$$

$$= \left\{\sum_{\mathbf{J}_\delta} + \sum_{\mathbf{J}_\delta^c}\right\} b_{\mathbf{j}_{2+},\mathbf{K}_{2+}}(\mathbf{p}_{2+}) b_{\mathbf{j_1},\mathbf{K_1-1}}(p_1)\left[\frac{\partial f(j_1/K_1,\mathbf{j}_{2+}\oslash\mathbf{K}_{2+},h)}{\partial p_1} + \zeta(\delta',\mathbf{p},h)\right]$$

$$= \frac{\partial f(\mathbf{p},h)}{\partial p_1} + \left\{\sum_{\mathbf{J}_\delta} + \sum_{\mathbf{J}_\delta^c}\right\} b_{\mathbf{j}_{2+},\mathbf{K}_{2+}}(\mathbf{p}_{2+}) b_{\mathbf{j_1},\mathbf{K_1-1}}(p_1)\left[\frac{\partial f(j_1/K_1,\mathbf{j}_{2+}\oslash\mathbf{K}_{2+},h)}{\partial p_1} - \frac{\partial f(\mathbf{p},h)}{\partial p_1} + \zeta(\delta',\mathbf{p},h)\right].$$

On $\mathbf{J}_\delta$, the term above in square brackets is bounded by $\frac{\varepsilon}{6}$, which then also bounds the first sum, and on $\mathbf{J}_\delta^c$ this term is bounded by $5M/\delta$. The probability of $\mathbf{J}_\delta^c$ is bounded by $2\sum_{i=1}^n \exp(-\delta^2 K_i/2)$, so the second sum is bounded by $\frac{10M}{\delta}\cdot \sum_{i=1}^n \exp(-\delta^2 K_i/2) \le \frac{10M}{\delta}\cdot\sum_{i=1}^n \frac{1}{\delta^2 K_i/2}$. Then $\mathbf{K} \ge 192nM/\varepsilon\delta^3$ implies that the second sum is bounded by $20\varepsilon/192 < \varepsilon/6$. This establishes the approximation property $\left|\frac{\partial B_K(\mathbf{p},\beta_K(h))}{\partial p_i} - \frac{\partial f(\mathbf{p},h)}{\partial p_i} < \varepsilon/3\right|$ at each ($\mathbf{p}$,h) in A.

A final step to establish (i) and (ii) uniformly considers the open cover of neighborhoods where the results hold (with tolerance ε/2 rather than ε/3), extracts finite sub-coverings for the compact domains, and uses the minimum value of δ from these finite sub-coverings. By construction, $\beta_{\cdot\mathbf{K}}$ retains the properties of f with respect to h; hence, in particular, if f is Lipschitz in H, then so is $\beta_{\cdot\mathbf{K}}$. ■

The next results will establish uniform convergence of empirical expectations for a family of functions that encompasses the applications in this paper. These results are obtained as specializations of the general theory of stochastic convergence treated in Dudley (2014), Kosorak (2008), Pollard (1984), and van der Vaart and Wellner (1996), referred to hereafter as VW. Let Y denote a closed subset of $\mathbb{R}^n$, $\mathcal{Y}$ denote the Borel σ-field of subsets of Y, and F denote a probability on $\mathcal{Y}$. Define a family $\mathcal{F}$ of functions f:Y ⟶ ℝ that is contained in the Banach space $\mathcal{L}_1$(Y,$\mathcal{Y}$,F) and includes the constant function f(y) ≡ 1. We assume that the functions in $\mathcal{F}$ are bounded by an envelope function f* ∈ $\mathcal{L}_1$(Y,,F); i.e., f* ≥ |f| for f ∈ $\mathcal{F}$. Let Θ denote a compact subset of $\mathbb{R}^d$, with a bound α > $\max(1,\max_{\theta\in\Theta}\|\theta\|)$. Assume that the functions in $\mathcal{F}$ are indexed by θ ∈ Θ and are Lipschitz with respect to this index; specifically, $|f_{\theta'}(y) - f_{\theta''}(y)| \le \|\theta' - \theta''\|\cdot f^*(y) \le \alpha\cdot f^*(y)$. We will call $\mathcal{F}$ with the properties above a *Lipschitz-parametric family*.

Let F_T denote the empirical probability defined by T independent draws $\{y_1,\ldots,y_T\}$ from F; i.e., for A ∈ $\mathcal{Y}$, $F_T(A) = \frac{1}{T}\sum_{t=1}^T \mathbf{1}(y_t \in A)$. For f ∈ $\mathcal{L}_1$(Y,,F) and a probability Q on $\mathcal{Y}$, define $\mathbf{E}_Q f \equiv \int_Y f(y)Q(dy)$ and $\mathbf{E}_T f \equiv \frac{1}{T}\sum_{t=1}^T f(y_t)$. Define $\|f\|_Q = \mathbf{E}_Q|f|$, and note that $\|f\|_F$ is the norm of $\mathcal{L}_1$(Y,,F). A strong law of large numbers establishes that $\mathbf{E}_T f \underset{as}{\longrightarrow} \mathbf{E}_F f$ pointwise for each f ∈ $\mathcal{F}$ and for f*. We give conditions under which this convergence is uniform on $\mathcal{F}$.

A measure of the "density" or "complexity" of $\mathcal{F}$ is its *bracketing number* $N_{[]}(\gamma,\mathcal{F},Q)$, defined for γ > 0 and a probability Q on $\mathcal{Y}$, the minimum cardinality of a family $\mathcal{F}_\gamma \subseteq \mathcal{L}_1$(Y,$\mathcal{Y}$,F), not necessarily a subset of $\mathcal{F}$, such that for each f ∈ $\mathcal{F}$, there are f′,f″ ∈ $\mathcal{F}$ satisfying f′ ≥ f ≥ f″ and $\mathbf{E}_Q$(f′ – f″) < γ. A related measure of the complexity of $\mathcal{F}$ is its *covering number* N(γ,$\mathcal{F}$,Q), defined for γ > 0 and a probability Q on $\mathcal{Y}$ as the minimum cardinality of a family $\mathcal{F}_\gamma \subseteq \mathcal{L}_1$(Y,$\mathcal{Y}$,F), not necessarily a subset of $\mathcal{F}$, such that for each f ∈ $\mathcal{F}$, $\inf_{f'\in\mathcal{F}_\gamma} \mathbf{E}_Q|f' - f| < \gamma$. Obviously, N(γ,$\mathcal{F}$,Q) ≤

$N_{[]}(\gamma,\mathcal{F},Q)$. We will be interested in families of functions for which the bracketing or covering number is finite. The following result specializes VW Theorem 2.7.11:

Lemma A.2. Consider a Lipschitz-parametric family $\mathcal{F}$ and a positive constant M > 1. For each probability Q on $\mathcal{Y}$ such that $\mathbf{E}_Q f^* \le M$ and each $\gamma > 0$, $N_{[]}(\gamma,\mathcal{F},Q) \le 2 + 2(8\alpha M/\gamma)^d$.

Proof: Let J be the largest integer no greater than $8\alpha M/\gamma$, and $\mathbf{j} = (j_1,\dots,j_d)$ a vector of indices with $1 \le j_i \le J$ for each i. Consider the family of open balls of radius $\gamma/2M$ centered at $b_j = (-\alpha + j_1\gamma/4M,\dots,-\alpha+j_d\gamma/4M)$. This family covers $\Theta \subseteq [-\alpha,\alpha]^d$ and contains J^d elements. Discard the balls that do not intersect Θ. From each of the remainder, select a point $\theta_j \in \Theta$ and let $\mathcal{F}_\gamma$ denote the family of functions $\min(f_{\theta_j} + \frac{\gamma f^*}{2M}, f^*)$ and $\max(f_{\theta_j} - \frac{\gamma f^*}{2M}, -f^*)$ plus f* and –f*. Then, $\mathcal{F}_\gamma$ contains at most $2(1+J^d)$ functions. For θ in the ball containing θ_j, the Lipschitz condition gives $|f_{\theta_i}(y) - f_\theta(y)| \le \frac{\gamma}{2M} f^*(y)$, implying $f_{\theta_j}(y) + \frac{\gamma}{2M} f^*(y) - f(y) \ge 0 \ge f_{\theta_j}(y) - \frac{\gamma}{2M} f^*(y) - f(y)$. Then $\mathcal{F}_\gamma$ brackets $\mathcal{F}$, and $N_{[]}(\gamma,\mathcal{F},Q) \le 2 + 2(8\alpha M/\gamma)^d$. ■

Augment the Lipschitz-parametric family $\mathcal{F}$ with the countable family $\mathcal{F}^0 \equiv \cup_{k=1}^{\infty} \mathcal{F}_{1/k}$ of the approximating functions in Lemma A.2 at tolerances γ = 1/k for k = 1,2,… ; i.e., consider the family $\mathcal{F}^* \equiv \mathcal{F} \cup \mathcal{F}^0$. Then the bound on bracketing numbers that the lemma establishes for $\mathcal{F}$ also holds for $\mathcal{F}^*$, and $\mathcal{F}^0$ is dense in $\mathcal{F}^*$. Then, $\mathcal{F}^*$ is said to be *Q-measurable* for any probability Q on $\mathcal{Y}$ such that $\mathbf{E}_Q f^* \le M$; see VW, 2.2.3 and 2.2.4.

Theorem A.3. Consider a Lipschitz-parametric family $\mathcal{F} \subseteq \mathcal{L}_1(Y,,F)$ that has an envelope $f^* \in \mathcal{L}_1(Y,\mathcal{Y},F)$. For each tolerance γ ∈ (0,1), $\lim_{T\to\infty} \mathrm{Prob}(\sup_{T'\ge T} \sup_{f\in\mathcal{F}} |(\mathbf{E}_{T'} - \mathbf{E})f| > \gamma) = 0$.

Proof: From the discussion following Lemma A.2, consider the augmented family $\mathcal{F}^*$ that contains $\mathcal{F}$ and also contains the countable dense subfamily $\mathcal{F}^0$. Given γ ∈ (0,1), the condition $\mathbf{E}_F f^* < \infty$ implies there exists a constant $M > \mathbf{E}_F f^*$ such that $\mathbf{E}_F f^*\cdot 1(f^*>M) < \gamma/4$. Define $\mathcal{F}^M = \{\min(M,\max(f,-M)) \mid f \in \mathcal{F}^*\}$. From Lemma A.2, the bracketing number bound established on $\mathcal{F}$ by the functions in $\mathcal{F}_\gamma$ also holds for $\mathcal{F}^M$ and the corresponding finite family $\mathcal{F}_\gamma^M = \{\min(M,\max(f,-M)) \mid f \in \mathcal{F}_\gamma\}$ for all probabilities Q on $\mathcal{Y}$, since $f_M^* = \min(M,\max(f^*,-M))$ is an envelope function for $\mathcal{F}^M$ and $\mathbf{E}_Q f_M^* \le M$. Then from Lemma A.2, $N(\gamma, \mathcal{F}^M, F_T) \le 2+2(8\alpha M/\gamma)^d$. This bound is independent of T. Then, the result follows for $\mathcal{F}^*$, and hence for $\mathcal{F}$, from VW Theorem 2.4.3. ■

The following result is stated in a form sufficient for our needs; for more general results, see VW, 2.6.17.

Theorem A.4. Consider a finite-dimensional linear subspace $\mathcal{K}$ of $\mathcal{L}_1(Y,,F)$. Without loss of generality, assume that $\mathcal{K}$ includes the function $f(y) \equiv 1$. For a fixed integer J, define $\mathcal{F}$ to be a subset of the family of functions of the form $\min(f_1,\dots,f_j)$ for $f_j \in \mathcal{K}, 1 \le j \le J$. Let $\mathcal{I}$ denote the family of indicator functions $i(y) = \mathbf{1}(f(y)>0)$ for $f \in \mathcal{F}$, and $\mathcal{G}$ denote the family of functions $g = f\cdot i$ for $f \in \mathcal{F}$ and $i \in \mathcal{I}$. Suppose $\mathcal{F}$ has an envelope function $f^* \in \mathcal{L}_1(Y,,F)$. Then, for each tolerance γ ∈ (0,1),

$$\lim_{T\to\infty} \mathrm{Prob}(\sup_{T'\ge T} \sup_{i\in\mathcal{I}} |(\mathbf{E}_{T'} - \mathbf{E})i| > \gamma) = 0 \text{ and } \lim_{T\to\infty} \mathrm{Prob}(\sup_{T'\ge T} \sup_{g\in\mathcal{G}} |(\mathbf{E}_{T'} - \mathbf{E})g| > \gamma) = 0.$$

Proof: The proof utilizes a geometric measure of the complexity of a family of functions $\mathcal{F}$ or a family of sets $\mathcal{C}$, the *Vapnik-Červonenkis (VC) index*, denoted $V(\mathcal{F})$ or $V(\mathcal{C})$; see VW 2.6.1, Dudley (2014, 2.6.1). Classes of functions or sets with a finite VC index are termed *VC-classes*. VW Lemma 2.6.15 establishes that $\mathcal{K}$ is a VC-class. Then VW, Lemma 2.6.18(ii) establishes that $\mathcal{F}$ is a VC-class with index $V(\mathcal{F})$, and the truncated class $\mathcal{F}^M = \{\min(M,\max(f,-M)) \mid f \in \mathcal{F}\}$ for $M > 0$ is a VC-class with index at most $V(\mathcal{F})+2$; see Dudley (2014, Theorem 4.41). Pollard (1984, Lemma 2.4.18) establishes that the family $\mathcal{C}$ of sets $C = \{y \in Y \mid f(y) > 0\}$ for $f \in \mathcal{F}$ is a VC-class, implying that $\mathcal{J}$ is a VC-class (see VW, p. 151, #9). VW Theorem 2.6.7 applied to $\mathcal{F}^M$ with envelope $f^* \equiv M$ or to $\mathcal{J}$ with envelope $i^* \equiv 1$ implies bounds $N(\gamma, \mathcal{F}^M, Q) \le K(M/\gamma)^{V(\mathcal{F})+2}$ and $N(\gamma, \mathcal{J}, Q) \le K(1/\gamma)^{V(\mathcal{F})+2}$ for $\gamma \in (0,1)$ and any probability Q on $\mathcal{Y}$, where K is a constant that does not depend on γ or M. Let $\mathcal{F}^M_{\gamma/2}$ and $\mathcal{J}_{\gamma/2M}$ denote the sets of centers of open balls of radius $\gamma/2$ and $\gamma/2M$ that cover $\mathcal{F}^M$ and $\mathcal{J}$ respectively, and satisfy $\text{card}(\mathcal{F}^M_{\gamma/2}) \le 2K(2M/\gamma)^{V(\mathcal{F})+2}$ and $\text{card}(\mathcal{J}_{\gamma/2M}) \le 2K(2M/\gamma)^{V(\mathcal{F})+2}$. Let $\mathcal{G}^M = \{i\cdot f \mid i \in \mathcal{J} \text{ and } f \in \mathcal{F}^M\}$. For $i \in \mathcal{J}$ and $f \in \mathcal{F}_M$, one has $\min_{i' \in \mathcal{J}_{\gamma/2M}} \min_{f' \in \mathcal{F}_{\gamma/2}} \mathbf{E}|i\cdot f - i'\cdot f'| \le \max_{f \in \mathcal{F}_M} \min_{i' \in \mathcal{J}_{\gamma/2M}} \mathbf{E}|(i - i')\cdot f| + \max_{i' \in \mathcal{J}_{\gamma/2M}} \min_{f' \in \mathcal{F}_{\gamma/2}} \mathbf{E}|(f - f')\cdot i'| < \gamma$. Then the covering number $N(\gamma, \mathcal{G}^M, Q)$ for any probability Q on $\mathcal{Y}$ is bounded by the number of functions in $\mathcal{G}_\gamma = \{i\cdot f \mid i \in \mathcal{J}_{\gamma/2M} \text{ and } f \in \mathcal{F}_{\gamma/2}\}$, which is in turn bounded by $4K^2(2M/\gamma)^{2V(\mathcal{F})+4}$. The countable families $\bigcup_{k=1}^{\infty} \mathcal{F}^M_{1/2K}$ and $\bigcup_{k=1}^{\infty} \mathcal{J}_{1/2kM}$ are dense in $\mathcal{F}^M$ and $\mathcal{J}$ respectively, so that these families are F-measurable. Then VW Theorem 2.4.3 applies to give the result. ■

Appendix B: Properties of Extreme Value Type 1 Random Variables

a. A standard Extreme Value Type 1 (EV1) random variable has CDF $F(\varepsilon) \equiv \exp(-e^{-\varepsilon})$, density $e^{-\varepsilon}\cdot\exp(-e^{-\varepsilon})$, and for $t < 1$ the moment generating function $\Gamma(1-t)$. Johnson and Kotz (1970, Ch. 21) show the linear transformation $\xi = v + \sigma\varepsilon$ with $\sigma > 0$ has CDF $\exp(-e^{-(\xi-v)/\sigma})$, mean $v + \sigma\gamma_0$, where $\gamma_0 = 0.5772156649\cdots$ is Euler's constant, median $v - \sigma \ln\ln 2$, mode v, and variance $\sigma^2\pi^2/6$. For $0 < \rho < 0.08$, the tails of $F(\varepsilon)$ satisfy $F(2\cdot\ln\rho) + 1 - F(-2\cdot\ln\rho) < \rho$ and $\int_{|\varepsilon|>-2\cdot\ln\rho} |\varepsilon| F(d\varepsilon) < \rho$. Also, $\mathbf{E}|\varepsilon| \le 1.219384$ (i.e., integrating by parts, $\mathbf{E}|\varepsilon| = \int_0^\infty \exp(-e^{\varepsilon})\,d\varepsilon + \int_0^\infty [1 - \exp(-e^{-\varepsilon})]d\varepsilon \le E_1(1) + \int_0^\infty \exp(-\varepsilon)\,d\varepsilon$, where $E_1(c) \equiv \int_c^\infty \frac{e^{-y}}{y}\,dy$ is the exponential integral with values given in Abramovitz and Stegum, 1964, Table 5.1). Finally, $\mathbf{E}\varepsilon^2 = \gamma_0^2 + \pi^2/6 = 1.978112\cdots$.

b. Consider $\mathbf{J} = \{0,\dots,J\}$, constants a_j and independent standard EV1 random variates ε_j for $j \in \mathbf{J}$, and a non-empty subset $\mathbf{C}$ of $\mathbf{J}$. Define $q_{\mathbf{C}} = \ln \sum_{j\in\mathbf{C}} e^{a_j}$ and $\xi_{\mathbf{C}} = \max_{j\in\mathbf{C}}(a_j + \varepsilon_j) - q_C$. Then $\xi_{\mathbf{C}}$ is again a standard EV1 random variable; i.e., $\text{Prob}(\xi_{\mathbf{C}} < c) = \text{Prob}(\varepsilon_j < c + q_{\mathbf{C}} - a_j \text{ for } j \in \mathbf{C}) = \prod_{j\in\mathbf{C}} \exp(-e^{-c-q_C+a_j}) \equiv \exp(-e^{-c})$. Given $k \in \mathbf{C}$, the probability of the event $Y_{\mathbf{C}}(k) = \{\varepsilon \mid a_k + \varepsilon_k \ge a_j + \varepsilon_j \text{ for } j \in C\}$ is multinomial logit,

$$L_{\mathbf{C}}(k) = \int_{\varepsilon_k=-\infty}^{+\infty} f(\varepsilon_k) \prod_{j\in\mathbf{C}\backslash\{k\}} F(\varepsilon_k + a_k - a_j)d\varepsilon_k = \int_{\varepsilon_k=-\infty}^{+\infty} e^{-\varepsilon_k}\exp(-e^{-\varepsilon_k}\textstyle\sum_{j\in\mathbf{C}} e^{a_j-a_k})d\varepsilon_k = \frac{e^{a_k}}{\sum_{j\in\mathbf{C}} e^{a_j}},$$

and for $\mathbf{A} \subseteq \mathbf{C}$, $L_{\mathbf{C}}(\mathbf{A}) = \frac{\sum_{j\in\mathbf{A}} e^{a_j}}{\sum_{j\in\mathbf{C}} e^{a_j}}$. The conditional CDF of ε_k, given $k \in \mathbf{C}$ and $Y_{\mathbf{C}}(k)$, is

$$\text{Prob}(\varepsilon_k < c \mid Y_{\mathbf{C}}(k)) = \frac{1}{L_{\mathbf{C}}(k)} \int_{\varepsilon_k=-\infty}^{c} f(\varepsilon_k) \prod_{j\in\mathbf{C}\backslash\{k\}} F(\varepsilon_k + a_k - a_j)d\varepsilon_k$$

$$= \frac{1}{L_{\mathbf{C}}(k)} \int_{\varepsilon_k=-\infty}^{c} e^{-\varepsilon_k}\exp(-e^{-\varepsilon_k}\textstyle\sum_{j\in\mathbf{C}} e^{a_j-a_k})d\varepsilon_k = \exp(-e^{-(c+a_k-q_{\mathbf{C}})}) \equiv F(c + a_k - q_{\mathbf{C}}).$$

Then the payoff $a_k + \varepsilon_k$, conditioned on the event $Y_{\mathbf{C}}(k)$, has the same CDF as $\xi_C + q_C$, and is therefore the same for all k. Term this the *Optimizer Invariance Property* (OIP). An immediate implication of OIP is

$$\gamma_0 + q_{\mathbf{C}} = \mathbf{E}(\xi_{\mathbf{C}} + q_{\mathbf{C}}) \equiv \mathbf{E}\max_{j\in\mathbf{C}}(a_j + \varepsilon_j) \equiv \mathbf{E}\{ a_k + \varepsilon_k \mid Y_{\mathbf{C}}(k)\}.$$

so these unconditional and conditional means are the same. This result is obtained in Dubin and McFadden (1984), Anas and Feng (1988), Resnick and Roy (1990), and Dubin (1985). A consequence is that if **B** and **C** are disjoint non-empty subsets of **J**, then the conditional (on $Y_{\mathbf{C}}(k)$ for some $k \in \mathbf{C}$) and unconditional expectations of utility differences are given by the same log sum difference:

$$\mathbf{E}\left\{\max_{j\in\mathbf{B}}(a_j + \varepsilon_j) - \max_{j\in\mathbf{C}}(a_j + \varepsilon_j) \mid Y_C(k)\right\} \equiv \mathbf{E}\{\max_{j\in\mathbf{B}}(a_j + \varepsilon_j) - \max_{j\in\mathbf{C}}(a_j + \varepsilon_j)\} \equiv \ln\frac{\sum_{j\in\mathbf{B}} e^{a_j}}{\sum_{j\in\mathbf{C}} e^{a_j}} .$$

If $k \notin \mathbf{C}$, then the conditional CDF of $\xi_{\mathbf{C}}$, given $a_k + \varepsilon_k > \xi_{\mathbf{C}} + q_{\mathbf{C}}$, is

$$\text{Prob}(\xi_{\mathbf{C}} < c \mid a_k + \varepsilon_k > \xi_C + q_C) = \frac{1}{L_{\mathbf{C}\cup\{k\}}(k)}\int_{\xi_{\mathbf{C}}=-\infty}^{c} f(\xi_{\mathbf{C}})[1 - F(\xi_C + q_{\mathbf{C}} - a_k)]d\xi_{\mathbf{C}}$$

$$= \frac{F(c)}{L_{\mathbf{C}\cup\{k\}}(k)} - \frac{\int_{\xi_{\mathbf{C}}=-\infty}^{c} e^{-\xi_C}\exp(-e^{-\xi_C}[1+e^{a_k-q_C}])d\xi_C}{L_{\mathbf{C}\cup\{k\}}(k)} = \frac{F(c) - L_{\mathbf{C}\cup\{k\}}(\mathbf{C})F(c + q_{\mathbf{C}} - \ln[e^{q_C}+e^{a_k}])}{L_{\mathbf{C}\cup\{k\}}(k)}.$$

Using the OIP, this result is unchanged if instead of a single alternative $k \notin \mathbf{C}$, there is a set of alternatives **A** with $\mathbf{A}\cap\mathbf{C} = \emptyset$ and either k maximizes $a_j + \varepsilon_j$ for $j \in \mathbf{A}$, with conditioning on the event $Y_{\mathbf{A}}(k)$, or $q_{\mathbf{A}} = \ln\sum_{j\in\mathbf{A}} e^{a_j}$ replaces a_k, a standard EV1 variate $\xi_{\mathbf{A}}$ replaces ε_k, and **A** replaces {k}.

Next, given $k \notin \mathbf{C}$, the conditional CDF of $\xi_{\mathbf{C}}$, given $a_k + \varepsilon_k < \xi_{\mathbf{C}} + q_{\mathbf{C}}$, is

$$\text{Prob}(\xi_C < c \mid a_k + \varepsilon_k < \xi_{\mathbf{C}} + q_{\mathbf{C}}) = \frac{1}{L_{\mathbf{C}\cup\{k\}}(\mathbf{C})}\int_{\xi_C=-\infty}^{c} f(\xi_C)F(\xi_{\mathbf{C}} + q_{\mathbf{C}} - a_k)d\xi_{\mathbf{C}}$$

$$= \frac{1}{L_{\mathbf{C}\cup\{k\}}(\mathbf{C})}\int_{\xi_{\mathbf{C}}=-\infty}^{c} e^{-\xi_C}\exp(-e^{-\xi_C}[1 + e^{a_k-q_C}])d\xi_{\mathbf{C}} \quad = F(c + q_{\mathbf{C}} - \ln[e^{q_C} + e^{a_k}]).$$

Again by the OIP, this result is unchanged if $k \in \mathbf{B}$ with $\mathbf{B}\cap\mathbf{C} = \emptyset$, $q_{\mathbf{B}} = \ln\sum_{j\in\mathbf{B}} e^{a_j}$ replaces a_k, a standard EV1 variate $\xi_{\mathbf{B}}$ replaces ε_k, and **A** replaces {k}.

c. Let **A**, **B**, **C** denote disjoint non-empty subsets of **J**. Define $q_{\mathbf{A}} = \ln\sum_{j\in\mathbf{A}} e^{a_j}$, and define $q_{\mathbf{B}}$ and $q_{\mathbf{C}}$ analogously. Define $\xi_{\mathbf{A}} = \max_{j\in\mathbf{A}}(a_j + \varepsilon_j) - q_{\mathbf{A}}$, with analogous definitions for $\xi_{\mathbf{B}}$ and $\xi_{\mathbf{C}}$, and let "**ABC**" denote the event $\xi_{\mathbf{A}} + q_{\mathbf{A}} > \xi_{\mathbf{B}} + q_{\mathbf{B}} > \xi_{\mathbf{C}} + q_{\mathbf{C}}$, and so on. The possible events and outcomes are given below:

	ABC	**ACB**	**BAC**	**BCA**	**CAB**	**CBA**
Choice at *a*	**A**	**A**	**A**	**C**	**C**	**C**
Choice at *b*	**B**	**C**	**B**	**B**	**C**	**C**
Type	difference	difference	compound	compound	compound	compound

The probability of the event **ABC** is

P(**ABC**) = $\int_{\xi_{\mathbf{B}}=-\infty}^{+\infty} f(\xi_{\mathrm{B}})\mathrm{F}(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}}-\mathrm{q}_{\mathbf{C}})[1-\mathrm{F}(\xi_{\mathrm{B}}+\mathrm{q}_{\mathbf{B}}-\mathrm{q}_{\mathbf{A}})]\mathrm{d}\xi_{\mathbf{B}}$

= $\int_{\xi_{\mathbf{B}}=-\infty}^{+\infty} \mathrm{e}^{-\xi_{\mathbf{B}}} \exp\left(-\mathrm{e}^{-\xi_{\mathbf{B}}}\left[1+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}-\mathrm{q}_{\mathbf{B}}}\right]\right) \mathrm{d}\xi_{\mathbf{B}} - \int_{\xi_{\mathbf{B}}=-\infty}^{+\infty} \mathrm{e}^{-\xi_{\mathbf{B}}} \exp\left(-\mathrm{e}^{-\xi_{\mathbf{B}}}\left[1+\mathrm{e}^{\mathrm{q}_{\mathbf{A}}-\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}-\mathrm{q}_{\mathbf{B}}}\right]\right) \mathrm{d}\xi_{\mathbf{B}}$

= P(**B**|**B**,**C**) – P(**B**|**A**,**B**,**C**) ≡ P(**B**|**B**,**C**)·P(**A**|**A**,**B**,**C**),

where P(**A**|**A**,**B**,**C**) = $\mathrm{e}^{\mathrm{q}_{\mathbf{A}}}/(\mathrm{e}^{\mathrm{q}_{\mathbf{A}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})$ and P(**B**|**B**,**C**) = $\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}/(\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})$. This formula gives the probability of any other of the events by substituting the corresponding permutation of **A**, **B**, **C**. Next,

E{$(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}})$·**1**(**ABC**)} = $\int_{\xi_{\mathbf{B}}=-\infty}^{+\infty}(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}}) f(\xi_{\mathbf{B}})\mathrm{F}(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}}-\mathrm{q}_{\mathbf{C}})[1-\mathrm{F}(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}}-\mathrm{q}_{\mathbf{A}})]\mathrm{d}\xi_{\mathbf{B}}$

= $\int_{\xi_{\mathbf{B}}=-\infty}^{+\infty}(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}})\mathrm{e}^{-\xi_{\mathbf{B}}}\left\{\exp\left(-\mathrm{e}^{-\xi_{\mathbf{B}}}\left[1+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}-\mathrm{q}_{\mathbf{B}}}\right]\right)-\exp\left(-\mathrm{e}^{-\xi_{\mathbf{B}}}\left[1+\mathrm{e}^{\mathrm{q}_{\mathbf{A}}-\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}-\mathrm{q}_{\mathbf{B}}}\right]\right)\right\}\mathrm{d}\xi_{\mathbf{B}}$

= $\{\mathrm{P}(\mathbf{B}|\mathbf{B},\mathbf{C})[\gamma_0+\ln(\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})]-\mathrm{P}(\mathbf{B}|\mathbf{A},\mathbf{B},\mathbf{C})[\gamma_0+\ln(\mathrm{e}^{\mathrm{q}_{\mathbf{A}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})]\}$,

The event **ACB** also has an expectation satisfying this "difference type" formula with **B** and **C** interchanged.

The event **BAC** has

E{$(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}})$ ·**1**(**BAC**)} = $\int_{\xi_{\mathbf{B}}=-\infty}^{+\infty}(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}})\mathrm{f}(\xi_{\mathbf{B}})\int_{\xi_{\mathbf{A}}=-\infty}^{\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}}-\mathrm{q}_{\mathbf{A}}}\mathrm{f}(\xi_{\mathbf{A}})\mathrm{F}(\xi_{\mathbf{A}}+\mathrm{q}_{\mathbf{A}}-\mathrm{q}_{\mathbf{C}})\mathrm{d}\xi_{\mathbf{A}}\mathrm{d}\xi_{\mathbf{B}}$

= $\int_{\xi_{\mathbf{B}}=-\infty}^{+\infty}(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}})\mathrm{f}(\xi_{\mathbf{B}})\int_{\xi_{\mathbf{A}}=-\infty}^{\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}}-\mathrm{q}_{\mathbf{A}}}\mathrm{e}^{-\xi_{\mathbf{A}}}\exp(-\mathrm{e}^{-\xi_{\mathbf{A}}}[1+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}-\mathrm{q}_{\mathbf{A}}}])\mathrm{d}\xi_{\mathbf{A}}\mathrm{d}\xi_{\mathbf{B}}$

= P(**A**|**A**,**C**) $\int_{\xi_{\mathbf{B}}=-\infty}^{+\infty}(\xi_{\mathrm{B}}+\mathrm{q}_{\mathbf{B}})\mathrm{e}^{-\xi_{\mathbf{B}}}\exp\left(-\mathrm{e}^{-\xi_{\mathbf{B}}}\right)\exp\left(-\mathrm{e}^{-\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{A}}-\mathrm{q}_{\mathbf{B}}}\left[1+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}-\mathrm{q}_{\mathbf{A}}}\right]\right)\mathrm{d}\xi_{\mathbf{B}}$

= P(**A**|**A**,**C**) $\int_{\xi_{\mathbf{B}}=-\infty}^{+\infty}(\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}})\mathrm{e}^{-\xi_{\mathbf{B}}}\exp\left(-\mathrm{e}^{-\xi_{\mathbf{B}}}\left[1+\mathrm{e}^{\mathrm{q}_{\mathbf{A}}-\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}-\mathrm{q}_{\mathbf{B}}}\right]\right)\mathrm{d}\xi_{\mathbf{B}}$

= P(**A**|**A**,**C**)P(**B**|**A**,**B**,**C**){ $\gamma_0+\ln(\mathrm{e}^{\mathrm{q}_{\mathbf{A}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})$} .

The events **BCA**, **CAB**, and **CBA** also have expectations satisfying this "compound type" formula with the corresponding permutations of **A**, **B**, and **C**.

Consider the event **AC**. From (b), **E** max($\xi_{\mathbf{A}}+\mathrm{q}_{\mathbf{A}}, \xi_{\mathbf{C}}+\mathrm{q}_{\mathbf{C}}$)]·**1**(**AC**) = $[\gamma_0\ +\ln(\mathrm{e}^{\mathrm{q}_{\mathbf{A}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})]\cdot\mathrm{P}(\mathbf{A}|\mathbf{A},\mathbf{C})$. Then,

E{[max($\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}}, \xi_{\mathbf{C}}+\mathrm{q}_{\mathbf{C}}$) – max($\xi_{\mathbf{A}}+\mathrm{q}_{\mathbf{A}}, \xi_{\mathbf{C}}+\mathrm{q}_{\mathbf{C}}$)]}·**1**(**AC**)

= **E**{[$\xi_{\mathbf{B}}+\frac{\mathrm{q}_{\mathbf{B}}}{\sigma}$ – ($\xi_{\mathbf{A}}+\mathrm{q}_{\mathbf{A}}$)]·**1**(**ABC**) + **E**{[$\xi_{\mathbf{C}}+\mathrm{q}_{\mathbf{C}}$ – ($\xi_{\mathbf{A}}+\mathrm{q}_{\mathbf{A}}$)]·**1**(**ACB**) + **E**{[[$\xi_{\mathbf{B}}+\mathrm{q}_{\mathbf{B}}$ – ($\xi_{\mathbf{A}}+\mathrm{q}_{\mathbf{A}}$)]]}·**1**(**BAC**)

= $\{\mathrm{P}(\mathbf{B}|\mathbf{B},\mathbf{C})[\gamma_0+\ln(\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})]-\mathrm{P}(\mathbf{B}|\mathbf{A},\mathbf{B},\mathbf{C})[\gamma_0+\ln(\mathrm{e}^{\mathrm{q}_{\mathbf{A}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})]\}$

+ $\{\mathrm{P}(\mathbf{C}|\mathbf{B},\mathbf{C})[\gamma_0+\ln(\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})]-\mathrm{P}(\mathbf{C}|\mathbf{A},\mathbf{B},\mathbf{C})[\gamma_0+\ln(\mathrm{e}^{\mathrm{q}_{\mathbf{A}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{B}}}+\mathrm{e}^{\mathrm{q}_{\mathbf{C}}})]\}$

$+ \mathrm{P}(\mathbf{A}|\mathbf{A},\mathbf{C})\mathrm{P}(\mathbf{B}|\mathbf{A},\mathbf{B},\mathbf{C})\{\gamma_0 + \ln(e^{q_A} + e^{q_B} + e^{q_C})\} - [\gamma_0 + \ln(e^{q_A} + e^{q_C})]\cdot\mathrm{P}(\mathbf{A}|\mathbf{A},\mathbf{C})$

$= -\mathrm{P}(\mathbf{C}|\mathbf{A},\mathbf{C})\cdot \ln(e^{q_A} + e^{q_B} + e^{q_C}) + \ln(e^{q_B} + e^{q_C}) - \mathrm{P}(\mathbf{A}|\mathbf{A},\mathbf{C})\cdot\ln(e^{q_A} + e^{q_C})$

$= \mathrm{P}(\mathbf{A}|\mathbf{A},\mathbf{C})\cdot\ln\frac{e^{q_B}+e^{q_C}}{e^{q_A}+e^{q_C}} + \mathrm{P}(\mathbf{C}|\mathbf{A},\mathbf{C})\cdot\ln\frac{e^{q_B}+e^{q_C}}{e^{q_A}+e^{q_B}+e^{q_C}}$.

Hence,

$$\mathbf{E}\{[\max(\xi_B + q_B, \xi_C + q_C) - \max(\xi_A + q_A, \xi_C + q_C)] \mid \mathbf{AC}\} = \ln\frac{e^{q_B}+e^{q_C}}{e^{q_A}+e^{q_C}} + \frac{\mathrm{P}(\mathbf{C}|\mathbf{A},\mathbf{C})}{\mathrm{P}(\mathbf{A}|\mathbf{A},\mathbf{C})}\cdot\ln\frac{e^{q_B}+e^{q_C}}{e^{q_A}+e^{q_B}+e^{q_C}}.$$

The first term in the last expression coincides with the unconditional expectation of the maximum, and the final term adjusts for the conditioning event. The adjustment is negative so that the information that the best in **A** is better than the best in **C** decreases the expected maximum utility over **B** and **C**. By application of the OIC as described at the end of (b), this result is the same no matter which event $Y_{\mathbf{A}}(k)$ occurs. Next,

$\mathbf{E}\{[\max(\xi_B + q_B, \xi_C + q_C) - \max(\xi_A + q_A, \xi_C + q_C)]\}\cdot\mathbf{1}(\mathbf{CA})$

$= \mathbf{E}\{[\xi_B + q_B - (\xi_C + q_C)]\cdot\mathbf{1}(\mathbf{BCA}) = \mathrm{P}(\mathbf{C}|\mathbf{A},\mathbf{C})\mathrm{P}(\mathbf{B}|\mathbf{A},\mathbf{B},\mathbf{C})[\gamma_0 + \ln(e^{q_A} + e^{q_B} + e^{q_C})]$

$- \mathrm{P}(\mathbf{C}|\mathbf{A},\mathbf{C})\cdot[\gamma_0 + \ln(e^{q_A} + e^{q_C})] + \mathrm{P}(\mathbf{C}|\mathbf{A},\mathbf{B},\mathbf{C})\cdot[\gamma_0 + \ln(e^{q_A} + e^{q_B} + e^{q_C})]$

$= \mathrm{P}(\mathbf{C}|\mathbf{A},\mathbf{C})\cdot\{\ln\frac{e^{q_B}+e^{q_C}}{e^{q_A}+e^{q_C}} - \ln\frac{e^{q_B}+e^{q_C}}{e^{q_A}+e^{q_B}+e^{q_C}}\}$,

and hence $\mathbf{E}\{\max(\xi_B + q_B, \xi_C + q_C) - \max(\xi_A + q_A, \xi_C + q_C) \mid \mathbf{CA}\} = \ln\frac{e^{q_B}+e^{q_C}}{e^{q_A}+e^{q_C}} - \ln\frac{e^{q_B}+e^{q_C}}{e^{q_A}+e^{q_B}+e^{q_C}}$.

As before, the first term in the last expression coincides with the unconditional expectation of the maximum, and the final term is a positive adjustment for the conditioning event, so that the information that the best in **C** is better than the best in **A** increases the expected maximum utility over **B** and **C**. Again, by application of the OIC, this result is the same no matter which event $Y_{\mathbf{A}}(k)$ occurs.

d. Now consider **J** = {0,…,J} and **C** = {0,…,J-1}. Assume that in a scenario change from m = *a* to m = *b*, constants $a_{jm} \equiv a_j$ for j ∈ **C** do not change, but $a_{Ja} \neq a_{Jb}$. Assume ε_j for j ∈ **J** is the same in both scenarios. Define $q_{\mathbf{C}} = \ln\sum_{j\in\mathbf{C}} e^{a_j}$ and $\xi_C = \max_{j\in\mathbf{C}}(a_j + \varepsilon_j) - q_{\mathbf{C}}$. There is an alternative k that maximizes $a_j + \varepsilon_j$ over j ∈ **C**, and from (b), the CDF of $a_k + \varepsilon_k$ given that k maximizes the payoff in **C** is the same as the CDF of $\xi_C + q_{\mathbf{C}}$. Define $\omega = \xi_C - \varepsilon_J$ and $L(w) \equiv \mathrm{Prob}(\omega \le w) = 1/(1+e^{-w})$. The possible events are then:

Event	Case	Condition	Probability	Payoff
Y_{bak}	$a_{Ja} < a_{Jb}$	$\xi_C+q_C < a_{Ja}+\varepsilon_J < a_{Jb}+\varepsilon_J$	$L(a_{Ja} - q_C)$	$a_{Jb} - a_{Ja}$
Y_{bka}	$a_{Ja} < a_{Jb}$	$a_{Ja}+\varepsilon_J < \xi_C+q_C < a_{Jb}+\varepsilon_J$	$L(a_{Jb} - q_C) - L(a_{Ja} - q_C)$	$a_{Jb} - q_C - \omega$
$\mathbf{Y}_{kba}$	$a_{Ja} < a_{Jb}$	$a_{Ja}+\varepsilon_J < a_{Jb}+\varepsilon_J < \xi_C+q_C$	$L(q_C - a_{Jb})$	0
$\mathbf{Y}_{abk}$	$a_{Ja} > a_{Jb}$	$\xi_C+q_C < a_{Jb}+\varepsilon_J < a_{Ja}+\varepsilon_J$	$L(a_{Jb} - q_C)$	$a_{Jb} - a_{Ja}$
Y_{akb}	$a_{Ja} > a_{Jb}$	$a_{Jb}+\varepsilon_J < \xi_C+q_C < a_{Ja}+\varepsilon_J$	$L(a_{Ja} - q_C) - L(a_{Jb} - q_C)$	$q_C - a_{Ja} + \omega$
Y_{kab}	$a_{Ja} > a_{Jb}$	$a_{Jb}+\varepsilon_J < a_{Ja}+\varepsilon_J < \xi_C+q_C$	$L(q_C - a_{Ja})$	0

Note that $\int_s^t \omega dL(\omega) = \omega L(\omega)]_s^t - \int_s^t \frac{e^\omega}{1+e^\omega} d\omega = tL(t) - sL(s) - \ln\frac{1+e^t}{1+e^s}$. Then the expected payoff in the event $\mathbf{Y}_{bka}$ is

$$\mathbf{E}\{(a_{Jb} - q_C - \omega)\cdot\mathbf{1}(\omega\in\mathbf{Y}_{bka})\} = (a_{Jb} - q_C)[L(a_{Jb} - q_C) - L(a_{Ja} - q_C)] - \int_{a_{Ja}-q_C}^{a_{Jb}-q_C} \omega dL(\omega)$$

$$= (a_{Jb} - q_C)[L(a_{Jb} - q_C) - L(a_{Ja} - q_C)] - (a_{Jb} - q_C)L(a_{Jb} - q_C) + (a_{Ja} - q_C)[L(a_{Ja} - q_C) + \ln\frac{1+e^{a_{Jb}-q_C}}{1+e^{a_{Ja}-q_C}}$$

$$= (a_{Ja} - a_{Jb})\, L(a_{Ja} - q_C) + \ln\frac{e^{q_C}+e^{a_{Jb}}}{e^{q_C}+e^{a_{Ja}}},$$

and the expected payoff in the event Y_{akb} is

$$\mathbf{E}\{(q_C - a_{Ja} + \omega)\cdot\mathbf{1}(\omega\in\mathbf{Y}_{bka})\} = (q_C - a_{Ja})[\, L(a_{Ja} - q_C) - L(a_{Jb} - q_C)] + \int_{a_{Jb}-q_C}^{a_{Ja}-q_C} \omega dL(\omega)$$

$$= (q_C - a_{Ja})[L(a_{Ja} - q_C) - L(a_{Jb} - q_C)] + (a_{Ja} - q_C)L(a_{Ja} - q_C) - (a_{Jb} - q_C)L(a_{Jb} - q_C) - \ln\frac{1+e^{a_{Ja}-q_C}}{1+e^{a_{Jb}-q_C}}$$

$$= (a_{Ja} - a_{Jb})\, L(a_{Jb} - q_C) + \ln\frac{e^{q_C}+e^{a_{Jb}}}{e^{q_C}+e^{a_{Ja}}}.$$

Combining these results with other payoffs in the table gives

Scenario *a* Choice	Case	Expected Payoff Given Choice
J	$a_{Ja} < a_{Jb}$	$a_{Jb} - a_{Ja}$
J	$a_{Ja} > a_{Jb}$	$\frac{1}{L(a_{Ja} - q_C)} \ln \frac{e^{q_C} + e^{a_{Jb}}}{e^{q_C} + e^{a_{Ja}}}$
k	$a_{Ja} < a_{Jb}$	$-\frac{L(a_{Ja} - q_C)}{L(q_C - a_{Ja})}(a_{Jb} - a_{Ja}) + \frac{1}{L(q_C - a_{Ja})} \ln \frac{e^{q_C} + e^{a_{Jb}}}{e^{q_C} + e^{a_{Ja}}}$
k	$a_{Ja} > a_{Jb}$	0

e. Assume scenarios m = *a*, *b*, a set of alternatives $\mathbf{J}_a = \mathbf{J}_b = \mathbf{J} = \{0,\dots,J\}$, and noise $\boldsymbol{\varepsilon}$ that is the same in both scenarios. Let a_{jm} denote constants. Order the alternatives so that $\Delta_i \equiv a_{ib} - a_{ia}$ is non-decreasing in i. Define non-decreasing constants $c_i = \Delta_i + a_{ka} - a_{rb}$; then $c_k = a_{kb} - a_{rb}$, and $c_r = a_{ka} - a_{ra}$. Let $\mathbf{A}_{jm}$ denote the event that alternative j is optimal in scenario m. Consider the event

$$\mathbf{B}_{kr} = \mathbf{A}_{ka}\cap\mathbf{A}_{rb} = \{\boldsymbol{\varepsilon} \mid \varepsilon_k + a_{ka} \ge \varepsilon_i + a_{ia} \text{ for } i \ne k \;\&\; \varepsilon_r + a_{rb} \ge \varepsilon_i + a_{ib} \text{ for } i \ne r\},$$

including both cases k = r and k ≠ r. The $\mathbf{B}_{kr}$ are disjoint for different k or for different r except for sets of probability zero, and satisfy $\mathbf{A}_{ka} = \cup_{r=0}^{J} \mathbf{B}_{kr}$ and $\mathbf{A}_{rb} = \cup_{k=0}^{J} \mathbf{B}_{kr}$. The event B_{kr} implies $(a_{rb} - a_{kb}) \ge \varepsilon_k - \varepsilon_r \ge (a_{ra} - a_{ka})$. Hence, B_{kr} is non-empty if and only if $a_{rb} - a_{ra} \ge a_{kb} - a_{ka}$, or equivalently $c_k \le c_r$, implying r ≥ k.

Consider the case r = k. Then for $i \neq k$, $\varepsilon_i \leq \varepsilon_k + a_{ka} - \max(a_{ia}, a_{ib} + a_{ka} - a_{kb})$. Then

$$P(B_{kk}) = \int_{\varepsilon_k=-\infty}^{+\infty} e^{-\varepsilon_k} \exp\left(-e^{-\varepsilon_k} \sum_{i=0}^{J} e^{-a_{ka}+\max(a_{ia},a_{ib}+a_{ka}-a_{kb})}\right) d\varepsilon_k = \frac{e^{a_{ka}}}{\sum_{i=0}^{J} e^{\max(a_{ia},a_{ib}+a_{ka}-a_{kb})}} .$$

Then, the conditional probability $P_{kb|ka}$ that the optimal choice in scenario *b* is k, given that the optimal choice in scenario *a* is k, satisfies

$$P_{kb|ka} \equiv P(B_{kk})/P(A_{ka}) = \frac{\sum_{i=0}^{J} e^{a_{ia}}}{\sum_{i=0}^{J} e^{\max(a_{ia},a_{ib}+a_{ka}-a_{kb})}} .$$

If r < k, B_{kr} is empty and $P_{rb|ka} = 0$. Finally, consider the case r > k. Then $\boldsymbol{\varepsilon} \in B_{kr}$ requires $\varepsilon_r + (a_{ra} - a_{ka}) \equiv \varepsilon_r - c_r \leq \varepsilon_k \leq \varepsilon_r + (a_{rb} - a_{kb}) \equiv \varepsilon_r - c_k$. Let $B_{kri} = \{\boldsymbol{\varepsilon} \in B_{kr} \mid \varepsilon_r - c_{i+1} \leq \varepsilon_k \leq \varepsilon_r - c_i\}$ for i = k,…,r-1 and consider the inequalities $\varepsilon_n \leq \min(\varepsilon_k + (a_{ka} - a_{na}), \varepsilon_r + (a_{rb} - a_{nb})) = (a_{ka} - a_{na}) + \min(\varepsilon_k, \varepsilon_r - c_n)$. If n > i, then $\varepsilon_r - c_n \leq \varepsilon_k$, implying $\varepsilon_n \leq (a_{ka} - a_{na} - c_n) + \varepsilon_r$; otherwise, $\varepsilon_n \leq (a_{ka} - a_{na}) + \varepsilon_k$. The probability of B_{kri} is then

$$P_{kri} = \int_{\varepsilon_r=-\infty}^{+\infty} e^{-\varepsilon_r} \exp\left(-e^{-\varepsilon_r} \sum_{n=i+1}^{J} e^{(a_{na}-a_{ka}+c_n)}\right) \int_{\varepsilon_k=\varepsilon_r-c_{i+1}}^{\varepsilon_r-c_i} e^{-\varepsilon_k} \exp\left(-e^{-\varepsilon_k} \sum_{n=0}^{i} e^{(a_{na}-a_{ka})}\right) d\varepsilon_k d\varepsilon_r$$

$$= \frac{e^{a_{ka}}}{\sum_{n=0}^{i} e^{a_{na}}} \left\{ \int_{\varepsilon_r=-\infty}^{+\infty} e^{-\varepsilon_r} \exp\left(-e^{-\varepsilon_r}\left[\sum_{n=i+1}^{J} e^{(a_{na}-a_{rb}+\Delta_n)} + \sum_{n=0}^{i} e^{a_{na}-a_{rb}+\Delta_i}\right]\right) d\varepsilon_r - \int_{\varepsilon_r=-\infty}^{+\infty} e^{-\varepsilon_r} \exp\left(-e^{-\varepsilon_r}\left[\sum_{n=i+1}^{J} e^{a_{na}-a_{rb}+\Delta_n} + \sum_{n=0}^{i} e^{a_{na}-a_{rb}+\Delta_{i+1}}\right]\right) d\varepsilon_r \right\}$$

$$= \frac{e^{a_{ka}}}{\sum_{n=0}^{i} e^{a_{na}}} \left\{ \begin{array}{l} \int_{\varepsilon_r=-\infty}^{+\infty} e^{-\varepsilon_r} \exp\left(-e^{-\varepsilon_r}\left[\sum_{n=0}^{J} e^{(a_{na}-a_{rb}+\max(\Delta_i,\Delta_n))}\right]\right) d\varepsilon_r \\ - \int_{\varepsilon_r=-\infty}^{+\infty} e^{-\varepsilon_r} \exp\left(-e^{-\varepsilon_r}\left[\sum_{n=0}^{J} e^{(a_{na}-a_{rb}+\max(\Delta_{i+1},\Delta_n))}\right]\right) d\varepsilon_r \end{array} \right\}$$

$$= \frac{e^{a_{ka}}}{\sum_{n=0}^{i} e^{a_{na}}} \left[\frac{e^{a_{rb}}}{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_i,\Delta_n)}} - \frac{e^{a_{rb}}}{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}\right] = \frac{e^{a_{ka}} e^{a_{rb}}}{e^{a_{na}+\max(\Delta_i,\Delta_n)}} \cdot \frac{(e^{\Delta_{i+1}} - e^{\Delta_i})}{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}} .$$

Then, the probability that the optimal choice in scenario *b* is r > k, given that the optimal choice in scenario *a* is k, satisfies $P_{rb|ka} \equiv \sum_{i=k}^{r-1} P_{kri}/P_{ka} = \sum_{i=k}^{r-1} \frac{e^{a_{rb}} \sum_{n=0}^{J} e^{a_{na}}}{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_i,\Delta_n)}} \cdot \frac{(e^{\Delta_{i+1}} - e^{\Delta_i})}{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}$.

Next, consider the expectations of $(a_{kb} + \varepsilon_k) \cdot \mathbf{1}(\boldsymbol{\varepsilon} \in B_{kk})$ for r = k and $(a_{rb} + \varepsilon_r) \cdot \mathbf{1}(\boldsymbol{\varepsilon} \in B_{kri})$ for k ≤ i < r, given $\boldsymbol{\varepsilon} \in A_{ka}$. First,

$$\mathbf{E}_{\boldsymbol{\varepsilon}|\mathbf{A_{ka}}}(a_{kb} + \varepsilon_k) \cdot \mathbf{1}(\boldsymbol{\varepsilon} \in B_{kk}) = \frac{1}{P_{ka}} \int_{\varepsilon_k=-\infty}^{+\infty} (a_{kb}+\varepsilon_k) e^{-\varepsilon_k} \exp\left(-e^{-\varepsilon_k} \sum_{i=0}^{J} e^{-a_{ka}+\max(a_{ia},a_{ib}+a_{ka}-a_{kb})}\right) d\varepsilon_k$$

$$= P_{kb|ka}\left[a_{kb} - a_{ka} + \ln\left(\sum_{i=0}^{J} e^{\max(a_{ia},a_{ib}+a_{ka}-a_{kb})}\right) + \gamma_0\right].$$

Second, for k ≤ i < r,

$$\mathbf{E}_{\boldsymbol{\varepsilon}|\mathbf{A_{ka}}}(a_{rb} + \varepsilon_r)\cdot\mathbf{1}(\boldsymbol{\varepsilon}\in B_{kri}) = \frac{P_{kri}}{P_{ka}}a_{rb}$$

$$+\frac{\sum_{n=0}^{J}e^{a_{na}}}{\sum_{n=0}^{i}e^{\,a_{na}}}\left\{\begin{array}{l}\int_{\varepsilon_r=-\infty}^{+\infty}\varepsilon_r e^{-\varepsilon_r}\exp\left(-e^{-\varepsilon_r}\left[\sum_{n=0}^{J}e^{a_{na}-a_{rb}+\max(\Delta_i,\Delta_n)}\right]\right)d\varepsilon_r \\ -\int_{\varepsilon_r=-\infty}^{+\infty}\varepsilon_r e^{-\varepsilon_r}\exp(-e^{-\varepsilon_r}\left[\sum_{n=0}^{J}e^{(a_{na}-a_{rb}+\max(\Delta_{i+1},\Delta_n))}\right])d\varepsilon_r\end{array}\right\}$$

$$=\frac{P_{kri}}{P_{ka}}\gamma_0 + \frac{\sum_{n=0}^{J}e^{a_{na}}\,e^{a_{rb}}}{\sum_{n=0}^{i}e^{\,a_{na}}}\left[\frac{1}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}}-\frac{1}{\sum_{n=0}^{J}e^{(a_{na}+\max(\Delta_{i+1},\Delta_n))}}\right]\cdot\ln\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}$$

$$+\frac{\sum_{n=0}^{J}e^{a_{na}}\,e^{a_{rb}}}{\sum_{n=0}^{i}e^{\,a_{na}}}\cdot\frac{1}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}\cdot\ \ln\frac{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}$$

$$=\frac{P_{kri}}{P_{ka}}\left[\gamma_0 + \ln\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}\right]+\frac{e^{a_{rb}}}{\sum_{n=0}^{i}e^{\,a_{na}}}\cdot\frac{\sum_{n=0}^{J}e^{a_{na}}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}\cdot\ \ln\frac{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}$$

Hence, the conditional expectation of $(a_{rb} + \varepsilon_r)\cdot\mathbf{1}(\boldsymbol{\varepsilon}\in B_{kr})$ given $\boldsymbol{\varepsilon}\in A_{ka}$ is

$$P_{rb|ka}\gamma_0 + \sum_{i=k}^{r-1}\frac{P_{kri}}{P_{ka}}\cdot\ln\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)} + \sum_{i=k}^{r-1}\frac{1}{\sum_{n=0}^{i}e^{\,a_{na}}}\cdot\frac{e^{a_{rb}}\cdot\sum_{n=0}^{J}e^{\,a_{na}}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}\cdot\ \ln\frac{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}.$$

Summing this expression over r > k gives

$$(1-P_{kb|ka})\gamma_0 + \sum_{r>k}\sum_{i=k}^{r-1}\frac{P_{kri}}{P_{ka}}\cdot\ln\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}$$

$$+\sum_{r>k}\sum_{i=k}^{r-1}\frac{1}{\sum_{n=0}^{i}e^{\,a_{na}}}\cdot\frac{e^{a_{rb}}\cdot\sum_{n=0}^{J}e^{\,a_{na}}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}\cdot\ \ln\frac{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}$$

$$= (1-P_{kb|ka})\gamma_0 + \sum_{i=k}^{J-1}\sum_{r=i+1}^{J}\frac{P_{kri}}{P_{ka}}\cdot\ln\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}$$

$$+\sum_{i=k}^{J-1}\frac{1}{\sum_{n=0}^{i}e^{\,a_{na}}}\cdot\frac{\sum_{r=i+1}^{J}e^{a_{rb}}\cdot\sum_{n=0}^{J}e^{\,a_{na}}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}\cdot\ \ln\frac{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}.$$

Combining this expression with the earlier conditional expectation for r = k,

$$\mathbf{E}_{\varepsilon|A_{ka}}\max_{r\geq k}(a_{rb} + \varepsilon_r) = \gamma_0 + \frac{\sum_{i=0}^{J}e^{a_{ia}}}{\sum_{i=0}^{J}e^{\max(a_{ia},a_{ib}+a_{ka}-a_{kb})}}\left[a_{kb}-a_{ka}+\ln\left(\sum_{i=0}^{J}e^{\max(a_{ia},a_{ib}+a_{ka}-a_{kb})}\right)\right]$$

$$+\sum_{i=k}^{J-1}\frac{\sum_{r=i+1}^{J}e^{a_{rb}}\sum_{n=0}^{J}e^{\,a_{na}}}{\sum_{n=0}^{J}e^{(a_{na}+\max(\Delta_i,\Delta_n))/\sigma}}\cdot\frac{(e^{\Delta_{i+1}}-e^{\Delta_i})}{\sum_{n=0}^{J}e^{(a_{na}+\max(\Delta_{i+1},\Delta_n))/\sigma}}\cdot\ln\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}$$

$$+\sum_{i=k}^{J-1}\frac{1}{\sum_{n=0}^{i}e^{\,a_{na}}}\cdot\frac{\sum_{r=i+1}^{J}e^{a_{rb}}\cdot\sum_{n=0}^{J}e^{\,a_{na}}}{\sum_{n=0}^{J}e^{(a_{na}+\max(\Delta_{i+1},\Delta_n))}}\cdot\ \ln\frac{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_i,\Delta_n)}}{\sum_{n=0}^{J}e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}\quad.$$

A consequence of this formula is

$$\mathbf{E}\ \max_{j\in\mathbf{J}}\left(a_{jb}+\varepsilon_j\right)|A_{ka}) - \mathbf{E}\ \max_{j\in\mathbf{J}}\left(a_{ja}+\varepsilon_j\right)|A_{ka})$$

$$= \frac{\sum_{i=0}^{J} e^{a_{ia}}}{\sum_{i=0}^{J} e^{\max(a_{ia}, a_{ib}+a_{ka}-a_{kb})}} \left[a_{kb} - a_{ka} + \sigma \cdot \ln\left(\sum_{i=0}^{J} e^{\max(a_{ia}, a_{ib}+a_{ka}-a_{kb})}\right)\right] - \ln \sum_{n=0}^{J} e^{a_{na}}$$

$$+ \sum_{i=k}^{J-1} \frac{\sum_{r=i+1}^{J} e^{a_{rb}} \sum_{n=0}^{J} e^{a_{na}}}{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_i,\Delta_n)}} \cdot \frac{(e^{\Delta_{i+1}} - e^{\Delta_i})}{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}} \cdot \ln \sum_{n=0}^{J} e^{a_{na}+\max(\Delta_i,\Delta_n)}$$

$$+ \sum_{i=k}^{J-1} \frac{1}{\sum_{n=0}^{i} e^{a_{na}}} \cdot \frac{\sum_{r=i+1}^{J} e^{a_{rb}} \cdot \sum_{n=0}^{J} e^{a_{na}}}{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}} \cdot \sigma \cdot \ln \frac{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_i,\Delta_n)}}{\sum_{n=0}^{J} e^{a_{na}+\max(\Delta_{i+1},\Delta_n)}}.$$

Appendix C. R-Code for the Discrete Welfare Calculus using a Synthetic Population

```
#Code to estimate losses from consumers not knowing their data were shared

sink('P:\\USER\\Kenneth.Train\\misperceptions\\simulation of privacy
results\\SimulationsForDan.txt')

#Attributes
# 1 price
# 2 Commercials shown between shows
# 3 Fast content availability
# 4 More TV shows
# 5 More movies
# 6 share usage data
# 7 share usage and personal data
# 8 No service

#Estimated parameters of distribution of WTP and scale
#alpha is scale parameter; 1/alpha is distributed log-normal
#WTPs are dsitributed normal
#Estimates are in Table 9 and 10 of Foundations

#Mean, stdev, and correlation matrix of underlying normals:
#Order: log(1/alpha), commercials, fast, mostly TV, mostly movies, share usage,
share personal and usage
normmn <- c(-2.0002, -1.562,  3.945,  -0.6988,  2.963, -0.6224, -2.705, -27.26)
normstd <- c(1.0637, 3.940, 3.631,  4.857,  2.524,  2.494,  6.751,  19.42)
r1 <- c(1, -0.5813, -0.1371,  0.0358, 0.0256, 0.0022, -0.1287, 0.2801)
r2 <- c(0,  1.0000,  0.1172, -0.3473, 0.0109, -0.2562, -0.0079, -0.4108)
r3 <- c(0,  0,       1.0000,  0.8042, -0.4019, -0.3542, -0.4206, 0.2391)
r4 <- c(0,  0,       0,       1.0000, -0.5890, -0.1695, -0.3328, 0.4616)
r5 <- c(0,  0,       0,       0,       1.0000,  0.5141, 0.5181, -0.0147)
r6 <- c(0,  0,       0,       0,       0,       1.0000, 0.9370, -0.0563)
r7 <- c(0,  0,       0,       0,       0,       0,      1.0000, -0.0975)
r8 <- c(0,  0,       0,       0,       0,       0,      0,       1.0000)

corrMat=rbind(r1,r2,r3,r4,r5,r6,r7,r8)
corrMat=corrMat+t(corrMat) - diag(1,8);


#Specification of services available and combinations of services.
# Look like netflix (N), Amazon Prime (A), huluplus (H), and combos eg NA

N <- c(7.99, 0, 0, 0, 1, 0, 0, 0)
```

```
A <- c(6.58, 0, 0, 1, 0, 0, 0, 0)
H <- c(7.99, 0, 1, 0, 0, 0, 0, 0)

nos <- c(0,0,0,0,0,0,0,1) #No service
#Create matrix of attributes of the 8 alternatives:
xmat <- rbind(N,A,H,N+A,N+H,A+H,N+A+H,nos)  #8 alts x 8 attributes
#Indicator of which alternatives have Hulu:
hasH <- c(0,0,1,0,1,1,1,0)
ndraws <- 1000000

samplen <- c(86, 14, 12, 35, 21, 3, 22, 107) #Number of people in survey who chose
each of the 8 alternatives

mktshares <- samplen/sum(samplen)
market <- sum(samplen)*(6/58) #in million. We know Hulu has 6m customers and 58
people in the survey have Hulu

#Create draws of coefficients
set.seed(1234)
coef <- matrix(rnorm(8*ndraws),8,ndraws)
coef <- matrix(rep(normmn,times=ndraws),8,ndraws) + diag(normstd) %*%
(t(chol(corrMat)) %*% coef)
print("Check mean, std, and correlation matrix of draws against true")
print("Means: simulated and true")
print(cbind(rowMeans(coef),normmn)) #Check against normmn
print("Stds: simulated and true")
print(cbind( sqrt(diag(cov(t(coef)))), normstd)) #Check against normstd
print("Correlation matrix, simulated first, then true")
print(cor(t(coef))) #Check against corrMat
print(corrMat)
wtpsharing <- coef[7,]
pcoef <- exp(coef[1,]); #For lognormally distributed price
coef[1,]<- -pcoef;  #First coef is for price
coef[2:8,] <- matrix(rep(pcoef, each=7),7,ndraws) * coef[2:8,] #Attribute coefs are
wtp times price coef

#Calculate representative decision utility and choice probabilities
u <- xmat %*% coef
eu <- exp(u)
eu[is.infinite(eu)] <- 10^300
p <- eu / matrix(rep(colSums(eu),each=8),8,ndraws)
s <- rowMeans(p)

# Adjust constants to equal market shares

alpha <- matrix(0,8,1)
oldu <- u

for(count in 1:20){
  alpha <- alpha+log(mktshares / s)
  u <- oldu+matrix(rep(alpha,times=ndraws),8,ndraws)
  eu <- exp(u)
  eu[is.infinite(eu)] <- 10^300
  p <- eu / matrix(rep(colSums(eu),each=8),8,ndraws)
  s <- rowMeans(p)
}
```

```
print("ASCs")
print(alpha)
print("Predicted and actual market shares at ASCs")
print(cbind(s,mktshares))


#Calculate welfare impact of lack of knowlegde of sharing by hulu-like service
#Actual attributes; which includes sharing of usage and personal data by hulu
#Same as above but now 1 in column 7 for Hulu, to indicate that Hulu shares
personal and usage info:

H[7] <- 1

xmatnew <- rbind(N,A,H,N+A,N+H,A+H,N+A+H, nos)  #8 alts x 8 attributes
unew <- xmatnew %*% coef
unew <- unew + matrix(rep(alpha,times=ndraws),8,ndraws)
eunew <- exp(unew);
eunew[is.infinite(eunew)] <- 10^300
pnew <- eunew / matrix(rep(colSums(eunew),each=8),8,ndraws)
newshares <- rowMeans(pnew)
diffu <- unew-u

hold <- log(colSums(eu)) / pcoef
lsdecision <- mean(hold)  #expected log sum based on decision utility
hold <- log(colSums(eunew)) / pcoef
lsrealized <- mean(hold) #expected log sum based on realized utility
hold <- colSums(p * diffu) / pcoef
squareloss<- mean(hold) #expected difference between perceived and actual utility
in money metric

hulusubscribers= t(mktshares) %*% hasH

print("Difference between peoples realized utility and decision utility for chosen
alternative")
print("in money metric.")
print("Aggregate, and per-person who subscribed to Hulu")
print(cbind((squareloss * market), (squareloss / hulusubscribers)))

print("Note: Conditional mean WTP (second number above) differs from unconditional
mean of 2.70.")

print("Difference between peoples realized utility and the utility they would have
obtained if informed");
print("in money metric.");
print("Aggregate, and per-person who subscribed Hulu")
print(cbind(((lsdecision-lsrealized+squareloss) *market),((lsdecision-
lsrealized+squareloss) / hulusubscribers)))
print("Hulu share")
print("Actual choices, informed choices, percent difference")
print(cbind(hulusubscribers,(t(newshares) %*% hasH),((t(mktshares-newshares) %*%
hasH) / hulusubscribers)))


#Break down analysis further by conditioning on choice and whether person likes or
dislikes sharing info.
```

```
poswtp <- coef[7,]>=0 #These people like sharing their information; others dislike
it.

everrors <- matrix(runif(8*ndraws),8,ndraws)
everrors <- -log(-log(everrors))
util <- u+everrors
util <- util / matrix(rep(pcoef,each=8),8,ndraws)  #So utils are back in money
metric
utilnew <- unew+everrors
utilnew <- utilnew / matrix(rep(pcoef,each=8),8,ndraws)

i=max.col(t(util))
inew=max.col(t(utilnew))
c=matrix(0,1,ndraws)
exper_util=matrix(0,1,ndraws)
cnew=matrix(0,1,ndraws)

for(n in 1:ndraws) {
   chosenalt <- i[n]
   c[1,n] <- util[chosenalt,n]
   exper_util[1,n] <- utilnew[chosenalt,n]
   newchosenalt <- inew[n]
   cnew[1,n] <- utilnew[newchosenalt,n]
}
Huluer <- i == 3 | i == 5 | i== 6 | i==7
NumHuluer <- sum(Huluer)
NumHuluerPosWTP <- sum(Huluer * poswtp)
NumHuluerNegWTP <- sum(Huluer *(1-poswtp))
print("Dollar difference in welfare relative what expected")
print("Aggregate, per person, per Hulu subscriber")
print("Everyone:")
xx <- exper_util-c
print(cbind(market * mean(xx), mean(xx), sum(xx)/NumHuluer))
print ("People whose dislike sharing:")
xx <-(exper_util-c) * (1-poswtp)
print(cbind(market * mean(xx), mean(xx), sum(xx)/NumHuluerNegWTP))
print("People whose like sharing:")
xx <-(exper_util-c) * poswtp
print(cbind(market * mean(xx), mean(xx), sum(xx)/NumHuluerPosWTP))

NumOther <- sum(1-Huluer)
NumOtherPosWTP <- sum(Huluer*poswtp)
NumOtherNegWTP <- sum(Huluer*(1-poswtp))
print("Dollar difference in welfare relative to being informed")
print("Aggregate, per person, average for Hulu subscribers, average for non-
subscribers")
print("Everyone:")
xx <- exper_util-cnew
print(cbind(market*mean(xx), mean(xx), mean(xx[Huluer==1]), mean(xx[Huluer==0]) ))
print("People whose dislike sharing:")
print(cbind(mean(1-poswtp)*market*mean(xx[poswtp==0]), mean(xx[poswtp==0]),
mean(xx[Huluer==1 & poswtp==0]), mean(xx[Huluer==0 & poswtp==0]) ))
print("People whose like sharing:")
print(cbind(mean(poswtp)*market*mean(xx[poswtp==1]), mean(xx[poswtp==1]),
mean(xx[Huluer==1 & poswtp==1]), mean(xx[Huluer==0 & poswtp==1]) ))
```

```
choicemat <- array(0,dim=c(8,8,2)) #Rows for actual choice, cols for choice if
informed, depth for wtp<>0
exper_diff <- array(0,dim=c(8,8,2))
welfare_diff <- array(0,dim=c(8,8,2))
for(rr in 1:8) {
  for(cc in 1:8) {
     k <- (i == rr) & (inew == cc) & (poswtp==0)
     choicemat[rr,cc,1] <- sum(k==1)
     exper_diff[rr,cc,1] <- sum(c[k==1]-exper_util[k==1])
     welfare_diff[rr,cc,1] <- sum(exper_util[k==1]-cnew[k==1])
     k <- (i == rr) & (inew == cc) & (poswtp==1)
     choicemat[rr,cc,2] <- sum(k==1)
     exper_diff[rr,cc,2] <- sum(c[k==1]-exper_util[k==1])
     welfare_diff[rr,cc,2] <- sum(exper_util[k==1]-cnew[k==1])
  }
}


print("Hulu subscriber or not")
matH <- matrix(c( hasH,(1-hasH)),8,2)
subscribe1 <- t(matH) %*% choicemat[,,1] %*% matH
subscribe2 <- t(matH) %*% choicemat[,,2] %*% matH
print("Share of population who chose row and would have chosen col")
print((subscribe1+subscribe2)/ndraws )
print("Of people who dislike sharing, share who chose row and would have chosen
col")
print( subscribe1/sum(subscribe1) )
print("Of people who like sharing, share who chose row and would have chosen col")
print( subscribe2/sum(subscribe2) )

sink()
```