Discrete Choice and Stochastic Utility Maximization

Ruud H. Koning and Geert Ridder

USC Center for Law, Economics & Organization
Research Paper No. C02-16

USC
THE LAW
SCHOOL

CENTER FOR LAW, ECONOMICS AND ORGANIZATION RESEARCH PAPER SERIES

Sponsored by the John M. Olin Foundation

University of Southern California Law School
Los Angeles, CA 90089-0071

Discrete Choice and Stochastic Utility Maximization

Ruud H. Koning[1]   Geert Ridder[2]

Keywords: discrete choice, stochastic utility maximization, nested multinomial logit model
JEL-classification: C25, D12

Summary Discrete choice models are usually derived from the assumption of random utility maximization. We consider the reverse problem, whether choice probabilities are consistent with maximization of random utilities. This leads to tests that consider the variation of these choice probabilities with the average utilities of the alternatives. By restricting the range of the average utilities we obtain a sequence of tests with fewer maintained hypotheses. In an empirical application, even the test with the fewest maintaned hypotheses rejects the hypothesis of random utility maximization.

1. Introduction

Consider an economic agent who must choose one of I alternatives that are indexed by $i = 1, \ldots, I$. We assume that the agent has ranked the I alternatives. This ranking can always be represented by a utility function $u_i, i = 1, \ldots, I$ with u_i the utility of alternative i and $u_i \geq u_j$ if and only if i is ranked at least as high as j. In this discrete choice problem, a rational agent chooses the most preferred alternative which is also the alternative that yields the highest level of utility.

Suppose that we observe the choice made by the agent. Can we test whether the agent has made a rational choice, *i.e.* has chosen the most preferred, utility-maximizing alternative? If we can observe or measure the utilities u_i, then testing for rational behaviour is straightforward. If we have no information on the utilities u_i, then any choice can be rationalized by an appropriate choice of the utilities, and only by observing repeated choices with identical utilities of the alternatives can we hope to discover irrational behaviour.

In this paper, we take an intermediate position, because we assume that we have some but limited knowledge of the utilities attached to the alternatives. We assume that this lack of knowledge can be adequately represented by letting the I-vector of utilities be a draw from an I-variate distribution, of which we know the mean. Hence, we can write

$$u_i = -v_i + \varepsilon_i, \; i = 1, \ldots, I A \tag{1}$$

where $-v_i$ is the known mean of u_i, and ε_i is a zero mean random variable. Expressing the mean as $-v_i$ simplifies conditions that involve cross-derivatives. It is also consistent

[1]Corresponding author, Department of Econometrics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands, email `r.h.koning@eco.rug.nl`.

[2]Department of Economics, Dept of Economics, Kaprielian Hall, University Park Campus, University of Southern California, Los Angeles, CA-90089, USA, email `gridder@usc.edu`.

with the special case that the mean utility is equal to $y - p_i$ with y income or total expenditure and p_i the price of alternative i (y does not affect the preference ordering and can be omitted). The joint distribution function of the I-vector ε is denoted by F. The model in equation (1) is the *(additive) random utility* model.

Imperfect knowledge of the utilities of the alternatives makes it harder to predict the behaviour of the agent. If we assume that the agent has a decision rule based on the utilities, we can derive *choice probabilities* $P_i(v), i = 1, \ldots, I$, that specify the probability that alternative i is chosen for a given decision rule and a given distribution of the utility levels. In particular, we can derive the choice probabilities on the assumption that the agent makes a rational choice, *i.e.* that he chooses the alternative with the highest level of utility.

If the vector of average utilities v is known, we can identify the choice probabilities $P_i(v), i = 1, \ldots, I$ by observing repeated choices of an individual if ε reflects intrapersonal variation in utilities, or by observing the choices of a group of individuals with the same value of v if ε reflects interpersonal variation in utilities. In applications, the number of individuals with the same value of v is small, because v is a function of the observed characteristics of the alternatives that may vary in the population[1]. If we specify v as a parametric function of these characteristics and also specify a parametric distribution for ε, a decision rule that depends solely on the random utilities gives an expression for the choice probabilities that is known up to a vector of parameters and can be estimated by MLE with data from a sample of the population. For instance, a multivariate normal distribution for ε together with the assumption of random utility maximization yields multinomial probit choice probabilities, and if the components of ε are i.i.d. with an extreme value distribution we obtain the multinomial logit model, again under the assumption of (random) utility maximization. Of the specifications that are routinely used in empirical applications only the nested multinomial logit model does not impose the assumption of (random) utility maximization.

Matzkin (1992, 1993) shows that if the assumption of random utility maximization is maintained and if the functions v satisfy certain conditions[2], the functions v and the distribution of ε are nonparametrically identified. Because we consider tests of the hypothesis of (random) utility maximization, we cannot assume that this hypothesis holds and Matzkin's results do not apply. Instead we assume that v is specified as a parametric function of the characteristics. For the choice probabilities $P_i(v), i = 1, \ldots, I$ we choose a flexible parametric function of v that is not consistent with (random) utility maximization for certain parameter values, e.g. the unrestricted nested multinomial logit model. An alternative is to consider the choice probabilities as a semi-parametric (multiple) index model and to estimate the parameters of v together with the functions $P_i, i = \ldots, I$. The tests proposed below depend on the specification of v, and a rejection of random utility maximization may be due to an incorrect specification of v. As far as we know, Matzkin's nonparametric estimator has not been used in empirical research. Current practice is to specify v as a parametric function of the characteristics. The estimated model is then

[1]Characteristics of the alternatives vary in the population, if their valuation depends on observed characteristics of the individuals. These observed individual characteristics that are the same in al alternatives may enter separately in v.

[2]She considers various restrictions, e.g. linear homogeneity in the subvector of the characteristics that vary over the alternatives. This generalizes the linear parametric specification that is used in most parametric models of the choice probabilities.

used e.g. to study the effect of the introduction of new alternatives. Because this relies on the assumption of stochastic utility maximization it is of interest to test this hypothesis. Rejection leads to caution in the use of the estimated model. If a respecified model passes the test, it increases the confidence in counterfactuals that are based on stochastic utility maximization.

Our tests are based on the choice probabilities as a function of v with v in some set $\mathcal{V} \subseteq \mathbb{R}^I$, i.e. we require that the choice probabilities satisfy a number of conditions on this set. In a finite sample of size T, we can obtain estimates $P_i(v_t), i =, \ldots, I, t = \ldots, T$, and this suggests that the smallest set $\mathcal{V}$ that we should consider is the set $v_t, t = 1, \ldots, T$. The other extreme is to choose $\mathcal{V} = \mathbb{R}^I$, but the requirement that the choice probabilities are consistent with stochastic utility maximization on the latter set may be too strong, if the v_t in the sample are in a small subset of $\mathbb{R}^I$. An intermediate case is to test whether the choice probabilities are consistent with utility maximization on a subset $\mathcal{V}$ that contains $v_t, t = 1, \ldots, T$, because on that subset the estimated choice probabilities are more reliable.

The compatibility problem has been studied before. First, there is an analogy with the classical problem of integrability of demand functions. McFadden (1981) explores this analogy which, of course, is not complete, because random utility models do not assume that agents have identical preferences. Second, there is a considerable literature on revealed stochastic preference. A book by Chipman, McFadden, and Richter (1990) surveys this field. An important result is the equivalence of stochastic utility maximization and the strong axiom of revealed stochastic preference (McFadden and Richter (1990)) that generalizes a classical result of Houthakker (1950) to random choice models. However, there is little overlap between our results and the results in the literature on revealed stochastic preference. The latter literature gives tests for rationality that apply, if the agent's choice is restricted to subsets of the I alternatives. Note the analogy with the classical integrability problem, where demands vary due to changes in prices and total expenditure (or level of utility), *i.e.* due to changes in the choice set. In our conditions for rational choice, agents choose on all occasions between all I alternatives. However, the I-variate distribution of utility levels differs, either over time between choices of a single agent, or over agents because of differences in the non-random components v.

Our results follow on a much smaller set of papers (Williams (1977), Daly and Zachary (1979), Börsch-Supan (1990)). The main advantage of our results is their relevance for econometric practice. In econometric applications there usually is no variation in the choice set, but there is variation in v. We are interested in necessary and sufficient conditions that can lead to econometric tests of the hypothesis of stochastic utility maximization. Although much work remains to be done, we believe that our conditions can be used to construct such tests.

The plan of the paper is as follows. In section 2 we study the additive random utility model of equation (1). Section 3 contains necessary and sufficient conditions for global compatibility. We also consider perfect aggregation of individual preferences, and ask whether the resulting representative agent model is helpful for studying the compatibility problem. In section 4 we derive conditions for local compatibility for two classes of sets of non-random components $\mathcal{V}$. Section 5 contains an application of the theory to choice of mode of payment. Section 6 concludes and gives some directions for future research. Because we try to be comprehensive, we include results that have appeared before. However, we provide new proofs for some of these results, so that the reader may find it worthwhile to reconsider known results.

2. Additive Random Utility, Stochastic Preferences, and Choice Probabilities

In equation (1) we have specified the basic Additive Random Utility Model (ARUM) that we use to represent the preferences. With respect to the joint distribution function of the random components we make the following assumptions.

Assumption 1 (independence). *The joint distribution function of ε does not depend on v for all $v \in \mathcal{V}$.*

Manski (1988b) considers estimation of binary choice models under the weaker assumption of median independence. Matzkin (1993) uses Manski's arguments to establish nonparametric identification of v as a function of characteristics of the alternatives that may vary over individuals, if the ε are independent and identically distributed with a distribution that depends on the characteristics. In that case the distribution of ε is not identified, and the identifiability of that distribution is essential for a test of stochastic utility maximization (Manski (1988a)).

In the ARUM the preference ordering is unchanged if we add a constant to all $v_i, i = , \ldots, I$. This suggests that if individuals choose the alternative with the highest utility, the choice probabilities $P_i(v), i = 1, \ldots, I$ should also be unchanged under this operation, i.e. they are *translation invariant*. If the independence assumption does not hold, then the choice choice probabilities do not necessarily have this property. For two alternatives the choice probability under utility maximization is $P_1(v) = H(v_2 - v_1|v)$ with H the cdf of $\varepsilon_2 - \varepsilon_1$ that depends on v, and only if the dependence is through $v_2 - v_1$ the choice probability is translation invariant. Failure of the translation invariance of the choice probabilities is evidence that the independence assumption does not hold.

The other assumptions on the distribution of ε are

Assumption 2 (absolute continuity). *The joint distribution of ε is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^I$. In other words, any (Borel-measurable) set of measure 0 according to Lebesgue measure is assigned measure 0 under F.*

Assumption 3 (non-defectiveness). *The joint distribution of ε is non-defective:*

$$\lim_{\varepsilon \to \infty} F(\varepsilon) = 1,$$
$$. \lim_{\varepsilon \to -\infty} F(\varepsilon) = 0$$

A consequence of the assumptions of absolute continuity and non-defectiveness[3] of F is, that

$$\Pr(u_i = u_j) = 0, \; i \neq j = 1, \ldots, I,$$

i.e. the probability of ties is 0.

[3] Assumption 3 can be weakened to $\Pr(\varepsilon_i = \infty, \varepsilon_j = \infty) = 0$ and $\Pr(\varepsilon_i = -\infty, \varepsilon_j = -\infty) = 0$ for $i \neq j = 1, \ldots, I$. Allowing *e.g.* $\Pr(\varepsilon_i = \infty) > 0$ implies that a fraction of the population will choose i even if v_i approaches $-\infty$. We exclude this possibility. In section 4 we shall see that the non-defectiveness assumption becomes empty if $\mathcal{V}$ is bounded.

An ARUM is defined as a random utility model of the form

$$u = -v + \varepsilon, \ v \in \mathcal{V}$$

with u, v, and ε I-vectors, and where the distribution function of ε, F, satisfies the independence, absolute continuity and non-defectiveness assumptions. Hence, an ARUM model is characterized by the triple $(I, F, \mathcal{V})$. The ARUM corresponds to a preference ordering over the I alternatives in the case that the agent considers attributes not observed by the econometrician or that his choice is genuinely random. Before we use it as the basic representation of such a preference ordering, we must verify that it does not impose restrictions on the preference ordering.

Consider an agent who must rank I alternatives. Without loss of generality, we can assume that he is not indifferent between any two alternatives. Hence, there are $I!$ possible complete rankings, and each ranking corresponds to a complete, transitive strict preference ordering R. Denote the set of all such strict preference orderings by $\mathcal{R}$. A *random preference* model assigns probabilities π_k to all $I!$ preference orderings in $\mathcal{R}$, *i.e.* it consists of a pair $(\mathcal{R}, \Pi)$, with Π the (discrete) probability distribution over the $I!$ preference orderings in $\mathcal{R}$. The random preference model is the most basic way to express limited knowledge of the preferences of agents. Hence, it is natural to ask whether the ARUM with a fixed v places restrictions on the random preference model. The answer is given in the following theorem.

Theorem 1. *Every ARUM with* $\mathcal{V} = \{v\}$ *implies a random preference model* $(\mathcal{R}, \Pi)$. *Conversely, every probability distribution* Π *over* $\mathcal{R}$ *can be represented by an ARUM with* $\mathcal{V} = \{v\}$ *for an appropriate choice of* F.

Proof. See Appendix A.

The attractive feature of random utility models is that they allow for variation in preferences. This variation is either interpersonal or intrapersonal. Psychological theories of choice (Thurstone (1927), Luce (1959)) concentrate on intrapersonal variation, *i.e.* they consider repeated choices by the same individual. Because the ε are assumed to represent idiosyncratic contributions to the utility levels, they are independent between choices. Econometric models of individual choice are usually estimated from cross-section data, in which one choice is observed for each of a number of individuals. In this situation it is most natural think of ε as interpersonal variation in preferences. Usually, the ε are assumed to be independent between individuals. We need repeated choices by a number of individuals to distinguish between the two forms of preference variation.

This paper takes the econometric point of view, and as a consequence, we shall concentrate on interpersonal variation in preferences. In the additive random utility model, this variation can be decomposed into variation in the observed utility components v, and variation in the unobserved utility components ε. The two sources of variation are assumed to be independent. The tests for rational choice that are discussed in the sequel exploit the existence of variation in v that is independent of variation in ε. In tests based on repeated choices by the same individual there is usually no variation in v. Instead, the individual is faced with restricted choice sets. The individual must choose from a subset of all I alternatives and this subset varies between choices (see *e.g.* McFadden and Richter (1990)). Although such tests are useful in experimental situations, where one has control over the choice set, they are less useful in econometric applications based on cross-section data. In most econometric studies all individuals face the same choice set. However, the observed utility components v, and, of course, the unobserved utility components ε, differ between individuals. In this situation the tests that are considered below apply.

A rational individual chooses the alternative that yields the highest level of utility. If preferences have the ARUM form, stochastic utility maximization implies that the choice probabilities are given by

$$\begin{aligned} &P_i(v) = \Pr(u_i > u_j, i \neq j = 1, \ldots, I) = \\ &\int_{-\infty}^{\infty} \frac{\partial F}{\partial \varepsilon_i}(\varepsilon_i - v_i + v_1, \ldots, \varepsilon_i, \ldots, \varepsilon_i - v_i + v_I)\mathrm{d}\varepsilon_i, \\ &i = 1, \ldots, I. A \end{aligned} \tag{2}$$

The last equality follows from the absolute continuity of F.

Although theorem 1 shows that the ARUM representation does not impose any restriction of the preference ordering, the assumptions 1-3 do impose restrictions. The rest of this section deals with the question whether these assumptions are needed. First we consider assumption 2.

It is well-known that models with a discrete dependent variable may be logically inconsistent. We refer to such a problem as a *coherency failure* (see *e.g.* Heckman (1978), Gourieroux, Laffont, and Montfort (1980)). A model is incoherent if the mapping from the unobservable random variables to the observable outcome variables is not well-defined, and as a consequence the sum of the probabilities of all outcomes is either strictly less than or greater than 1. In the case of maximization of ARUM preferences, the unobservables are ε, the observable outcome is the chosen alternative and the mapping of ε to $\{1, \ldots, I\}$ is the subscript of $\operatorname{argmax}(u_1, \ldots, u_I)$.

Now let $I = 2$, and let F be such that[4]:

$$\Pr(u_1 = u_2) > 0.$$

Then

$$P_1(v) + P_2(v) = \Pr(u_1 > u_2) + \Pr(u_2 > u_1) = 1 - \Pr(u_1 = u_2) < 1,$$

the obvious problem being that the mapping from unobservables to observables is not well-defined if $u_1 = u_2$. Hence, to prevent a coherency failure, the distribution of $\varepsilon_2 - \varepsilon_1$ has to be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. Because the marginal distribution of ε_1 is arbitrary, we can choose it to be absolutely continuous and hence F is absolutely continuous. We conclude that if the stochastic utility maximization model is coherent, then F can be chosen to be absolutely continuous[5].

We have proved the following theorem.

[4]This occurs if the distribution of $\varepsilon_2 - \varepsilon_1$ has an atom. It is equivalent to the inclusion of nonstrict preferences in $\mathcal{R}$.

[5]Note, however, that distributions exist that yield a coherent stochastic utility maximization model even though they are not absolutely continuous. For example, let ε_1 have a discrete distribution and let ε_2 have a continuous distribution. The distribution of $\varepsilon_2 - \varepsilon_1$ will be continuous and the resulting stochastic utility maximization model will be coherent even though the joint distribution of $(\,\varepsilon_1 \quad \varepsilon_2\,)'$ is not absolutely continuous.

Theorem 2. *Consider an individual who has preferences which are of the ARUM form. We assume that the individual is rational, i.e. that the observed choice is obtained by stochastic utility maximization. Then this stochastic utility maximization model is coherent if and only if the distribution of*

$$\begin{pmatrix} \varepsilon_2 - \varepsilon_1 & \ldots & \varepsilon_I - \varepsilon_1 \end{pmatrix}'$$

is absolutely continuous with respect to Lebesgue measure.

3. Global Compatibility with Stochastic Utility Maximization

3.1. Global Compatibility in Discrete Choice Models

In this section we present a simple derivation of the necessary and sufficient conditions for the compatibility of choice probabilities $P_i(v)$, $i = 1, \ldots, I$ with stochastic utility maximization. We shall assume that v can take any value in $\mathbb{R}^I$. The necessary and sufficient conditions were first given by Daly and Zachary (1979). The simple arguments given here are also helpful in understanding later sections of this paper.

First, we define *global compatibility* with stochastic utility maximization. Here, and in the rest of this paper, stochastic utility maximization means maximization of preferences of ARUM form.

Definition 1. *The set of choice probabilities $P_i(v)$, $i = 1, \ldots, I$ is globally compatible with stochastic utility maximization, if for all $v \in \mathbb{R}^I$ we can write*

$$P_i(v) = \Pr(\varepsilon_j - v_j < \varepsilon_i - v_i; i \neq j = 1, \ldots, I),\ i = 1, \ldots, I A \tag{3}$$

with ε a stochastic I-vector with a non-defective and absolutely continuous (with respect to the Lebesgue measure) distribution that does not depend on v.

The $(I-1)$-vectors v^i and ε^i are obtained from the I-vectors v and ε by deleting the i-th element. Daly and Zachary (1979) state that the following conditions are necessary and sufficient for global compatibility. If a condition is subscribed by i or i and j, it holds for all $i = 1, \ldots, I$ or $i \neq j = 1, \ldots, I$, respectively.

Necessary and sufficient conditions for global compatibility.
For all $v \in \mathbb{R}^I$:

$$P_i(v) \geq 0,\ \sum_{i=1}^{I} P_i(v) = 1\text{k} \tag{C-1}$$

$$P_i(v) = P_i(v + c\iota_I) \text{ for all } c \in \mathbb{R} \text{ (translation invariance)k} \tag{C-2}$$

$$\lim_{v_i \to -\infty} P_i(v) = 1,\ \lim_{v_j \to -\infty} P_i(v) = 0\text{k} \tag{C-3}$$

$$P_i(v)\text{k} \quad \text{is differentiable with respect to } v^i \text{ and}$$

$$\frac{\partial^{(I-1)} P_i}{(\partial v)^i}(v) \geq 0 \text{ (non-negativity)k} \tag{C-4}$$

$$\frac{\partial P_i}{\partial v_j}(v) = \frac{\partial P_j}{\partial v_i}(v) \text{ (symmetry).k} \tag{C-5}$$

In condition (C-4), as well as in the sequel, $\partial v_1 \cdots \partial v_{i-1} \partial v_{i+1} \cdots \partial v_I$ is abbreviated as $(\partial v)^i$. These conditions are known in the literature as the Daly–Zachary or Daly–Zachary–Williams conditions.

Theorem 3 (Daly and Zachary). *Conditions (C-1)–(C-5) are necessary and sufficient for global compatibility with stochastic utility maximization.*

Proof. See Appendix B.

The first condition states that all choice probabilities are non-negative and that some alternative is chosen. According to (C-2), only the differences in average utilities determine the choice probabilities, not the absolute levels. The third condition requires that an alternative is chosen with probability 1, if its utility increases without bound. Condition (C-4) states that if all alternatives, except the i-th, become less attractive, the probability of choosing the i-th alternative should not decrease. Finally, (C-5) is the discrete choice analogue of the symmetry condition in demand analysis. A comparison of the conditions (C-1)-(C-5) with those for the integrability of demand systems can be found in Appendix C.

Remark 1. The definition of global compatibility implies that there exists a stochastic I-vector ε that satisfies equation (3). The proof shows that the choice of ε is not unique. The choice probabilities determine h_1, and by equation (B-4) also $h_2, \ldots, h_I$. In other words, they determine the distributions of $\varepsilon^1 - \varepsilon_1 \iota_{I-1}, \ldots, \varepsilon^I - \varepsilon_I \iota_{I-1}$. One marginal distribution, *e.g.* the distribution of ε_1, can be chosen arbitrarily. From equation (B-4) it follows that any one of the h_i determines all the other h_i's. For $I = 2$ this expression reduces to

$$h_1(v_2 - v_1) = h_2(v_1 - v_2),$$

i.e., h_2 is obtained by reflection of h_1 around 0.

Remark 2. The original Daly and Zachary (1979) paper does not contain a proof. The only published proof is that by McFadden (1981) (see his Theorem 5.1, assertion 3, pp. 212–213, and the proof in the Appendix 5.23). McFadden uses the same construction of the distribution function of ε, *i.e.* using alternative 1 as a reference alternative (see (5.131), p. 263). This establishes directly that

$$P_1(v) = \Pr(\varepsilon^1 - \varepsilon_1 \iota_{I-1} < v^1 - v_1 \iota_{I-1}).A \tag{4}$$

Next, he proves that equation (4) holds for $i = 1, \ldots, I$, by showing that the choice probabilities can be obtained as minus the gradient with respect to v of a function, the Social Surplus function that, can be defined using F (see (5.132), p. 264). These derivatives have the form of equation (2). Implicitly, this establishes that the choice of 1 as a reference alternative is arbitrary. Our proof is more direct, because we need not establish the existence and differentiability of a Social Surplus function. Instead, we make direct use of the symmetry condition (C-5) to show that the choice of the reference alternative is arbitrary. Our method of proof can be easily adapted to derive conditions for local compatibility.

Remark 3. Using theorem 3 we can make a rather surprising observation. Assume that the population of agents consists of two sub-populations. Fraction p chooses an alternative by maximizing a (random) utility function. Fraction $1-p$ picks an alternative at random. The choice probabilities for the whole population are

$$P_i(v) = p\tilde{P}_i(v) + (1-p)\frac{1}{I},\ i = 1, \ldots, I,$$

where the $\tilde{P}_i(v)$ satisfy (C-1)–(C-5). Now note that the $P_i(v)$ also satisfy (C-1)–(C-5). Hence, although only a fraction p of the agents makes a rational choice, the population choice probabilities are compatible with stochastic utility maximization.

Remark 4. The (multinomial) logit and the (multinomial) probit models satisfy the conditions for global compatibility for all values of their parameters. The Nested Multinomial Logit model satisfies (C-1), (C-2), (C-3) and (C-5) for all parameter values and (C-4) only if the association parameter is in the $(0, 1]$ interval.

If we let v^{ikl} be a k-subvector of v^i, and $l = 1, \ldots, \binom{I-1}{k}$, it is not difficult to see that (C-4) can be replaced by

$$P_i(v) \text{ is differentiable with respect to } v^i \text{ and } \frac{\partial^k P_i}{\partial v^{ikl}}(v) \geq 0\text{k} \tag{C'-4}$$

Hence, we have

Corollary 1. *Conditions (C-1), (C-2), (C-3), (C'-4), and (C-5) are necessary and sufficient for global compatibility with stochastic utility maximization.*

The proof of theorem 3 contains a further useful corollary.

Corollary 2. *The choice probabilities $P_i(v)$, $i = 1, \ldots, I$ are globally compatible with stochastic utility maximization if and only if there exist density functions $h_1, \ldots, h_I$ on $\mathbb{R}^{(I-1)}$ that for all $v \in \mathbb{R}^I$, $i, j = 1, \ldots, I$ satisfy*

$$h_i(v^i - v_i\iota_{I-1}) = h_j(v^j - v_j\iota_{I-1}),\ i \neq jA \tag{5}$$

and

$$P_i(v) = \int_{-\infty}^{v^i - v_i\iota_{I-1}} h_i(w)dw.A \tag{6}$$

By a change of variables we obtain a third corollary.

Corollary 3. *The choice probabilities $P_i(v)$, $i = 1, \ldots, I$ are globally compatible with stochastic utility maximization if and only if there is a density function h_1 on $\mathbb{R}^{(I-1)}$ such that for all $v \in \mathbb{R}^I$ and $i = 2, \ldots, I$ we have*

$$P_1(v) = \int_{-\infty}^{v^1 - v_1\iota_{I-1}} h_1(w)dw,A \tag{7}$$

and

$$P_i(v) = \int_{v_i - v_1}^{\infty} \int_{-\infty}^{w_{i-1}+(v_2-v_1)-(v_i-v_1)} \cdots \int_{-\infty}^{w_{i-1}+(v_I-v_1)-(v_i-v_1)} h_1(w)dw_{I-1}\cdots dw_1 dw_{i-1}.A \tag{8}$$

This corollary implies that the specification of one density function is sufficient to determine all choice probabilities. This density function is the density function of the $(I-1)$-vector $(\varepsilon_2 - \varepsilon_1, \ldots, \varepsilon_I - \varepsilon_1)'$, where the first alternative is arbitrarily taken as a reference alternative.

4. Local Compatibility with Stochastic Utility Maximization

In the definition of global compatibility the observed utility components v can take any value in $\mathbb{R}^I$. In local compatibility v is restricted to a subset of $\mathbb{R}^I$. Of course, local compatibility is weaker than global compatibility.

Definition 2. *The set of choice probabilities $P_i(v)$, $i = 1, \ldots, I$ is locally compatible with stochastic utility maximization on a set $\mathcal{V} \subset \mathbb{R}^I$, if for $i = 1, \ldots, I$ and all $v \in \mathcal{V}$ we can write*

$$P_i(v) = \Pr(\varepsilon_j - v_j < \varepsilon_i - v_i;\ j = 1, \ldots, I,\ j \neq i)\text{k} \tag{9}$$

with ε a stochastic I-vector with a non-defective and absolutely continuous distribution that does not depend on v.

Local compatibility was introduced in Börsch-Supan (1990), although he does not give a formal definition of the concept. Local compatibility is closer to econometric practice than global compatibility. In practice, v does not take on all values in $\mathbb{R}^I$, but we usually have a finite number of observed v_t, $t = 1, \ldots, T$. We specify choice probabilities, and ask whether these choice probabilities are consistent with utility maximization on a set $\mathcal{V}$ with $v_t \in \mathcal{V}$ for $t = 1, \ldots, T$. In Börsch-Supan's study the choice probabilities are obtained by fitting a flexible parametric functional form, the Nested Multinomial Logit model (NMNL), to the observed v_t and the corresponding observed choices. If the association parameters of the NMNL model are outside the $(0, 1]$ interval, condition (C-4) is violated. The other conditions for compatibility are satisfied for all parameter values (Börsch-Supan (1990)). Hence, the NMNL model is not globally compatible if a dissimilarity parameter is outside the $(0, 1]$ interval.

Now choose $a, b \in \mathbb{R}^I$ such that $a \leq v_t \leq b$ for $t = 1, \ldots, T$. We can ask under which conditions the fitted choice probabilities are locally compatible with stochastic utility maximization on $\mathcal{V} = [a, b]$[6]. The following theorem gives necessary and sufficient conditions.

Theorem 4. *The choice probabilities $P_i(v), i = 1, \ldots, I$ are locally compatible with stochastic utility maximization on a bounded interval $\mathcal{V} = [a, b]$ if and only if for all $v \in [a, b]$ (C-1), (C-2), (C'-4) and (C-5) hold.*

Proof. See Koning and Ridder (1994).

Remark 5. This result is stronger than that obtained in Börsch-Supan (1990). First, we do not require that (C-1), (C-2), and (C-5) are globally satisfied. Second, the theorem

[6] Strictly, local compatibility on a *bounded* interval is not possible: if equation (9) holds for all $v \in [a, b]$ it holds for all v such that $v + c\iota_I \in [a, b]$ for some $c \in \mathbb{R}$. Hence, we must choose a normalization. For that purpose, we express v in deviation from v_1, and we take $\mathcal{V}$ as the set $\left\{v \in \mathbb{R}^I \mid v - v_1\iota_I \in [a, b]\right\}$. Because $a_1 = b_1 = 0$, it suffices to specify $\tilde{\mathcal{V}} = [a^1, b^1] \subset \mathbb{R}^{(I-1)}$.

gives necessary and sufficient conditions. Third, Börsch-Supan does not prove local compatibility as defined above. His suggested distribution of $(\,\varepsilon_2-\varepsilon_1 \quad \ldots \quad \varepsilon_I-\varepsilon_1\,)'$ is not absolutely continuous with respect to the Lebesgue measure, and he does not indicate how the resulting ties will be resolved. The proof uses the representation of the choice probabilities of corollary 3 of theorem 3. We find a density function h_1^* that satisfies this equation for all $v \in [a,b]$.

Remark 6. Condition (C-3) is not required for local compatibility. Since equation (9) only has to be satisfied on an interval, we have more freedom in choosing the distribution of ε. In particular, we can always choose it to be non-defective.

Remark 7. The conditions in theorem 4 and corollary 1 of theorem 3 are identical, except for (C-3).

Theorem 4 gives necessary and sufficient conditions for local compatibility on an interval. If the estimated $v_t, t=1,\ldots,T$ are bounded, this is a natural choice, in particular if the model is to be used to predict the effect of changes in the characteristics of the alternatives and/or the individuals on the choice probabilities. If the set $\mathcal{V}$ is not an interval, then the conditions in Theorem 4 are neither necessary, nor sufficient for random utility maximization.

We consider choice between two alternatives, *i.e.* $I=2$. Let us assume that the choice probabilities are translation invariant, and differentiable, *i.e.* (C-1), (C-2), and the first part of (C'-4) are satisfied. Then for $I=2$ (C-5) is also satisfied[7]. The choice probabilities $P_1(v)$ and $P_2(v)$ can be expressed as

$$\begin{aligned} P_1(v) &= \int_{-\infty}^{v_2-v_1} h_1(w)\mathrm{d}w, \\ P_2(v) &= \int_{v_2-v_1}^{\infty} h_1(w)\mathrm{d}w, \end{aligned}$$

with $h_1(w) = \frac{\partial P_1(0,w)}{\partial w}$. Let h_1 and P_1 be as in figure 1.

Note that h_1 is non-negative on $\tilde{\mathcal{V}}_1 = (-\infty, w_0] \cup [w_1, \infty)$ and hence (C'-4) is satisfied. However, it is not possible to find a density function h_1^* that coincides with h_1 on $\tilde{\mathcal{V}}_1$ and also satisfies

$$\begin{aligned} P_1(v) &= \int_{-\infty}^{v_2-v_1} h_1^*(w)\mathrm{d}w, \\ P_2(v) &= \int_{v_2-v_1}^{\infty} h_1^*(w)\mathrm{d}w, \end{aligned}$$

because $P_1(0,w_1) < P_1(0,w_0)$ and hence h_1^* has to be negative for some values of w. The conditions in Theorem 4 are hence not sufficient for compatibility on $\tilde{\mathcal{V}}_1$. It is also not necessary, because local compatibility holds on e.g. $\tilde{\mathcal{V}}_2$ and (C'-4) does not hold on this set.

[7]These assumptions apply if we fit a flexible functional form, that satisfies all conditions, except the non-negativity condition.

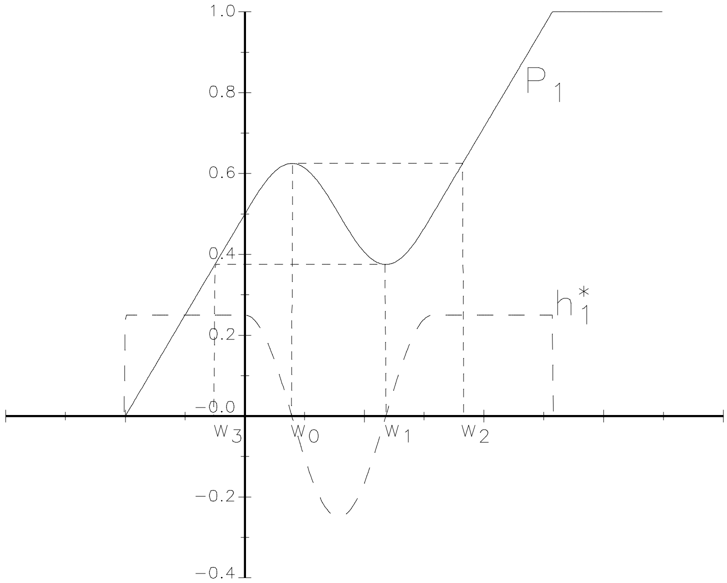

Figure 1. Compatibility for arbitrary intervals

The key is that nonnegativity (C'-4) on an interval implies monotonicity, but that is not true on a set that is the union of disjoint intervals as in the example. The assumption of monotonicity is fundamental, and in the example the choice probability is compatible with utility maximization for any set on which $P_1(0, w)$ is monotonous. If $\mathcal{V} = v_t, t = 1, \ldots, T$, then monotonicity is equivalent to

$$v_{2t} - v_{1t} \geq v_{2s} - v_{1s} \Rightarrow P_1(v_{1t}, v_{2t}) \geq P_1(v_{1s}, v_{2s}).A \tag{10}$$

Next we generalize monotonicity to choices between $I \geq 3$ alternatives.

Is it possible to find necessary and sufficient conditions for other choices of $\mathcal{V}$? We shall consider a choice of $\mathcal{V}$ that can be seen as the opposite extreme, namely a finite set of distinct points. This choice of $\mathcal{V}$ is of considerable practical interest, because in practice an econometrician has a finite sample $\mathcal{V} = \{v_1, \ldots, v_T\}$ of observed average utilities. If for every v_t he observes a large number of choices made by distinct agents, he can determine the corresponding choice probabilities $P_i(v_t)$, $i = 1, \ldots, I$. If the number of observed choices for each t is small, he can use either a local averaging method, *e.g.* a kernel estimate, or a flexible functional form, *e.g.* the NMNL model, to estimate $P_i(v_t)$, $i = 1, \ldots, I$, $t = 1, \ldots, T$. How can he decide whether these (estimated) choice probabilities are (locally) compatible with stochastic utility maximization (on $\mathcal{V}$)?

First, we consider the case with three alternatives ($I = 3$) with the first alternative as the reference alternative. The observed utilities of the first observation is shown as point $\tilde{w}_1$ in Figure 2. If the choices are made by stochastic utility maximization, then,

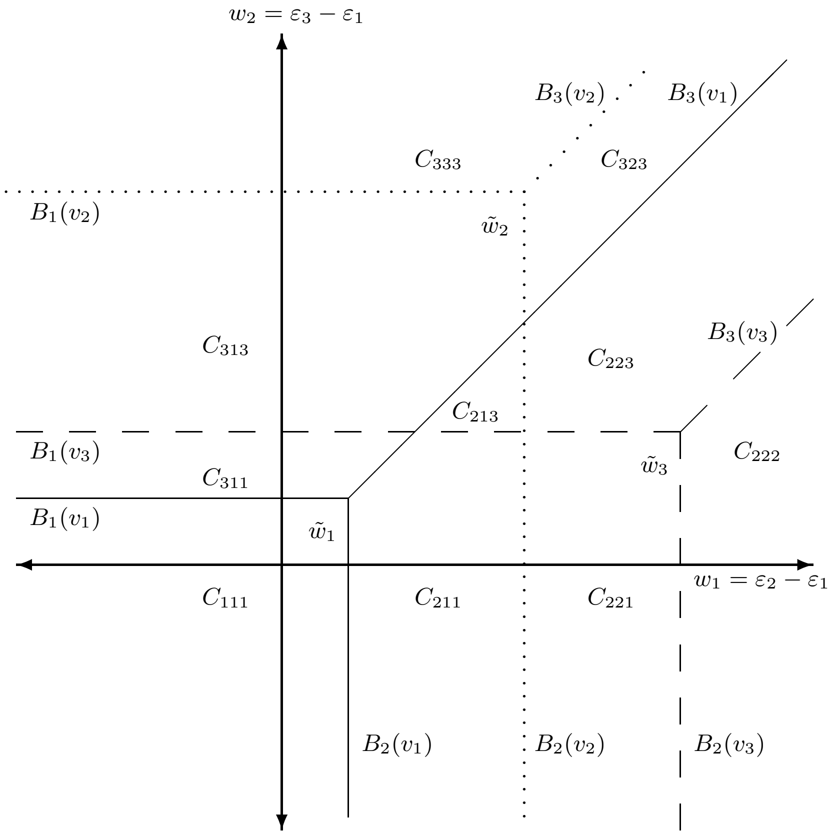

Figure 2. Choice probabilities and compatibility ($\tilde{w}_t = (v_{2t} - v_{1t},\, v_{3t} - v_{1t})$, $t = 1, 2, 3$).

according to corrolary 3, we can write the choice probabilities as

$$P_i(v) = \int_{B_i(v)} h_1(w)\mathrm{d}w, A \tag{11}$$

with

$$\begin{aligned}
B_1(v_1) &= \left\{w \in \mathbb{R}^2 \mid w_1 \leq v_2 - v_1, w_2 \leq v_3 - v_1\right\} \\
B_2(v_1) &= \left\{w \in \mathbb{R}^2 \mid w_1 > v_2 - v_1, w_2 \leq v_3 - v_1\right\} \\
B_3(v_1) &= \left\{w \in \mathbb{R}^2 \mid w_1 > v_2 - v_1, w_2 > v_3 - v_1\right\}.
\end{aligned}$$

The first observation induces a partition of $\mathbb{R}^2$ into three sets, and the three choice probabilities correspond to the integral of the the density function of the random terms over these three sets. If we add a second observation ($\tilde{w}_2$), we obtain a similar partition denoted by $B_1(v_2)$, $B_2(v_2)$, and $B_3(v_2)$. The third observation $\tilde{w}_3$ gives the partition $B_1(v_3)$, $B_2(v_3)$, and $B_3(v_3)$. The intersection of these three partitions is itself a partition of $\mathbb{R}^2$ and this partition consists of the sets C_{111} to C_{333} in Figure 2. The order of the indices of the sets C reflect the observation, and the index itself indicates to which B-set the set belongs, so for example $C_{213} = B_2(v_1) \cap B_1(v_2) \cap B_3(v_3)$. Hence, each set $B_i(v_t)$ can be written as the union of a collection of sets C, for example in Figure 2 we have $B_1(v_3) = C_{111} \cup C_{311} \cup C_{211} \cup C_{221}$. We denote the index set over which the union is taken by $J_i^*(v_t)$, so by definition $B_i(v_t) = \cup_{J_i^*(v_t)} C_{i_1 i_2 \cdots i_T}$. Now we can rewrite (11) for each observation and for each choice probability as

$$P_i(v) = \int_{B_i(v)} h_1(w)\mathrm{d}w = \sum_{J_i^*(v_t)} \int_{C_{i_1 i_2 \cdots i_T}} h_1(w)\mathrm{d}w.A \tag{12}$$

If we denote each term of the summation by $A_{i_1 i_2 \cdots i_T}$, so

$$A_{i_1 i_2 \cdots i_T} = \int_{C_{i_1 i_2 \cdots i_T}} h_1(w)\mathrm{d}w,$$

we are able to give the general result.

Theorem 5. *The choice probabilities $P_i(v), i = 1, \ldots, I$ are locally compatible with stochastic utility maximization on $\mathcal{V} = \{v_1, \ldots, v_T\}$ if and only if condition (C-1) holds on $\mathcal{V}$ and the system of equations in $A_{i_1 i_2 \cdots i_T}$*

$$P_i(v_t) = \sum_{(i_1, i_2, \ldots, i_T) \in J_i^*(v_t)} A_{i_1 i_2 \cdots i_T} \tag{13}$$

$t = 1, \ldots, T, A\ i = 1, \ldots, I$ *has a non-negative solution.*

Proof. See Appendix D.

In the example of Figure 2, the choice probabilities are compatible with stochastic utility maximization if the following set of equations has a non-negative solution:

$$P_1(v_1) = A_{111} \tag{14-a}$$
$$P_2(v_1) = A_{211} + A_{213} + A_{222} + A_{223} + A_{221} \tag{14-b}$$
$$P_3(v_1) = A_{311} + A_{313} + A_{333} + A_{323} \tag{14-c}$$
$$P_1(v_2) = A_{111} + A_{311} + A_{313} + A_{213} + A_{211} \tag{14-d}$$
$$P_2(v_2) = A_{221} + A_{223} + A_{323} + A_{222} \tag{14-e}$$
$$P_3(v_2) = A_{333} \tag{14-f}$$
$$P_1(v_3) = A_{111} + A_{311} + A_{211} + A_{221} \tag{14-g}$$
$$P_2(v_3) = A_{222} \tag{14-h}$$
$$P_3(v_1) = A_{313} + A_{333} + A_{213} + A_{223} + A_{323} \tag{14-i}$$

Remark 8. The existence of a solution to equation (13) implies that the choice probabilities are translation invariant. However, conditions (C-3) and (C-5) are not needed. If (C-4) holds for all v, then equation (13) has a non-negative solution.

Remark 9. The equation system in equation (13) has some resemblance with an equation system in McFadden and Richter (1990) (see (3.8), p. 172). However, as stressed before, any resemblance is superficial, because the rationality test of theorem 5 is different from the McFadden/Richter test. The McFadden/Richter test exploits variation in the choice probabilities if the choice set is restricted to a subset C_i of alternatives, while in the test of theorem 5 the agents always choose between I alternatives but with varying average utilities. In the notation of McFadden and Richter, we have $m = 1$ and $B = \{1, \ldots, I\}$. Hence, the only C_i that are used in the test have $C_i \subset B$ with B fixed. It is not difficult to see that in that case the McFadden/Richter test degenerates. It is always possible to solve their equation system if for all alternatives the probability that the alternative is most preferred in the (random) preference ordering is positive. For ARUM preferences (with fixed v) this condition is always satisfied.

As a corollary to theorem 5, we give a necessary condition on the choice probabilities that can be checked easily.

Corollary 4. *If the choice probabilities $P_i(v)$ are compatible with stochastic utility maximization, then for each pair (t, t'):*

$$B_i(v_t) \subseteq B_i(v_{t'}) \Rightarrow P_i(v_t) \leq P_i(v_{t'}).A \tag{15}$$

If $I = 3$, it is easily seen that each pair (t, t') yields two restrictions of the form (15) (see also figure 2). If $I = 2$, the condition of corollary 4 guarantees that the distribution function of $\varepsilon_2 - \varepsilon_1$ is non-decreasing in $v_{2t} - v_{1t}, \quad t = 1, \ldots, T$ and in that case (15) is also sufficient. This is not true if $I \geq 3$, as the following example illustrates[8].

In Figure 2 let the vectors of choice probabilities be $P(v_1) = (0, 0, 1)'$, $P(v_2) = (0, 1, 0)'$ and $P(v_3) = (1, 0, 0)'$, which satisfy condition (C-1). One can easily check that the observed choice probabilities satisfy the condition of corollary 4. However, the set of equations (14-a)–(14-i) does not have a non-negative solution for A. Hence, the condition of Theorem 5 is not satisfied.

The test of theorem 5 is not superfluous: in Appendix E we give an example of choice probabilities that satisfy the conditions of theorem 5, but that do not satisfy those of theorem 4.

If we compare theorems 3, 4, and 5, we can make a number of observations. First, globally compatible choice probabilities are locally compatible on any bounded interval and also locally compatible on every finite set. Choice probabilities that are locally compatible on an interval are also locally compatible on any finite subset of that interval. Second, the number of conditions diminishes if the 'measure' of $\mathcal{V}$ becomes smaller. In theorem 4 we do not need (C-3) and the condition (13) in theorem 5 is implied by the the non-negativity condition (C-4). The symmetry condition (C-5) is not needed for and is not implied by local compatibility on a finite set. Translation invariance is implicit in the condition of theorem 5. Third, if the 'measure' of $\mathcal{V}$ becomes smaller the freedom in choosing the distribution of ε increases. Hence, it becomes easier to satisfy the compatibility conditions. Finally, we have not considered the case that $\mathcal{V}$ is the union of a number of disjoint intervals. We conjecture that the necessary and sufficient conditions for that case are that the conditions of theorem 4 hold for every bounded interval and that for every subset of $\mathcal{V}$ the conditions of theorem 5 are satisfied.

[8]We owe this example to Jan Karel Lenstra.

5. An Application: Choice of Mode of Payment

In this section we apply the theory in a test of random utility maximization. When someone pays for an over-the-counter purchase, he has the choice between different modes of payment. We consider the choice between cash payment and payment by check. Our data are for the Netherlands in 1987 and at that time payment by credit card was still rare. The additive random utility model is

$$\nu_{it} = -v_{it} + \varepsilon_{it} = -\beta_i' x_t + \varepsilon_{it}.$$

The average utility v_{it} is a function of individual characteristics and the amount of the transaction and that function differs between the alternatives. The data is a sample of 225 transactions[9]. For each transaction the mode of payment is recorded, as well as the amount paid, size of the household, age of the head of the household, gender of the person making the transaction, and household income.

Let D be a dummy with value 0 if the payment was by cheque and 1 if cash was used. Under the assumption of random utility maximization, the probability of paying cash is

$$\Pr(D = 1|x) = \Pr(\varepsilon_1 - \varepsilon_2 \leq v_1 - v_2) = F(v_1 - v_2),A \tag{16}$$

We have $v_{1t} - v_{2t} = \beta' x_t$ with $\beta = \beta_1 - \beta_2$. Because D is a 0-1 variable we have

$$\mathcal{E}(D|x) = \Pr(D = 1) = G(\beta' x).A \tag{17}$$

and this is in the form of a single index model.

In our test of stochastic utility maximization we follow the example in Section 4. We maintain the assumptions that the choice probabilities are translation invariant and differentiable. A test of random utility maximization then amounts to a test of non-negativity, in this case test of whether F is non-decreasing. A test of global compatibility tests whether F is non-decreasing everywhere and a test of compatibility on an interval tests whether that is true on a specific interval, usually an interval that contains $\beta' x_t, t = 1, \ldots, T$. Finally, the weakest test is to check

$$\beta' x_t \geq \beta' x_s \Rightarrow F(\beta' x_t) \geq F(\beta' x_s)$$

There are a number of estimators for the parameters β of a single index model. These estimators estimate β up to scale. We use the average derivative estimator of Stoker (1991, 1992). This estimator is $\sqrt{n}$ consistent. In the second step we estimate F by a nonparametric (kernel) regression of D on $\hat{\beta}' x$. The kernel estimator does not impose that F is non-decreasing.

The average derivative estimates show that the probability of using cash for a transaction depends negatively on the amount of the transaction and positively on the age and household income of the individual who pays. The estimate of F is shown in Figure 3 for $h = 0.071$ and $h = 0.10$[10].

[9] The data are derived from the 1987 wave of a panel, the Intomart Bestedingen Index. The panel has slightly more than 1000 households. We restrict ourselves to households that have a bank account and therefore access to both payment alternatives. In case more than one payment has been recorded for one household we have randomly selected one. More details are given in Koning and Ridder (2000).

[10] The first choice of bandwidth follows the suggestion in Silverman (1986) for univariate kernel density estimation.

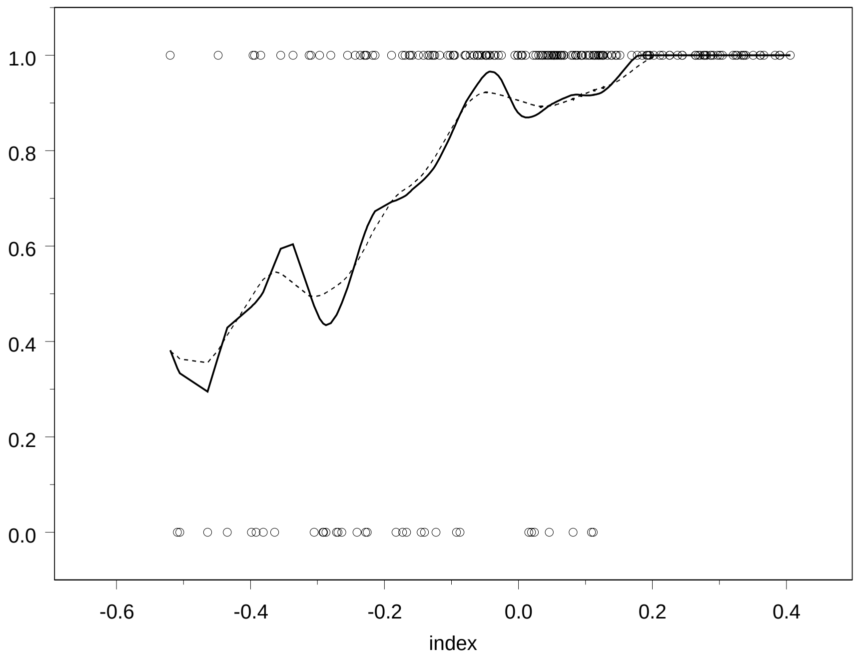

Figure 3. Nonparametric regression of choice (cash payment=1) on index, $h = 0.071$ (solid line) and $h = 0.10$ (dashed line).

The estimate $\hat{F}$ is not monotonic for both bandwidths. For sufficiently large h the regression curve is nondecreasing. A search shows that this is the case for $h \geq h_c = .15$. Bowman, Jones, and Gijbels (1998) have proposed a test for the hypothesis that a regression function is non-decreasing. In this test the estimated model is used to simulate D_t for all t. Using the new sample we reestimate β and F, the latter with the critical bandwidth $h_c = .15$. We check whether the estimated curve is nondecreasing. This is repeated a number of times (1000 in our test). The p-value of the test is the fraction non-monotonic curves in the (1000) simulated samples. With 1000 samples we overwhelmingly reject the null-hypothesis of monotonicity of the regression curve with $h = 0.071$: the p-value is 0.971. For the second curve ($h = 0.10$), the p-value is 0.944 so also for that curve monotonicity is decisively rejected.

One reason for the rejection could be that the function v is misspecified, as discussed in Section 1. To check this we also performed the test for a specification in which the squares of all variables are included in v. The p-value for this test is .99 and so we still reject the monotonicity of F.

As argued in Section 4, rejection of monotonicity of F precludes the existence of an interval on which the choice probability is compatible with stochastic utility maximization. Hence, in this case the rejection of global compatibility is a rejection of local compatibility on an interval.

Finally, we compute the fraction of pairs $\hat{\beta}'x_t, \hat{\beta}'x_{t+1}$ with $\hat{F}(\hat{\beta}'x_t) < \hat{F}(\hat{\beta}'x_{t+1})$. We have not developed the distribution theory for this test. The fraction can be positive for non-decreasing functions due to sampling variation in the estimated choice probabilities. In our example the fraction is 0.16 (37 out of 224).

6. Conclusion

The conditions in the theorems 3, 4, 5 can be used to construct rationality tests in discrete choice problems. The tests based on theorems 3 and 4 require the estimation of the choice probabilities for choices that are not observed in the data. In general, this is achieved by fitting a smooth and sufficiently flexible set of functions to the observed choices. The test of theorem 3 is completely based on the properties of these functions. The test of theorem 5 does not require extrapolation of the choice probabilities for average utilities that are not observed. This test has no maintained assumptions beyond those of the ARUM. In an empirical example we reject the hypothesis of random utility maximization, both globally and on an interval.

REFERENCES

Börsch-Supan, A. (1990). On the compatibility of nested logit models with utility maximization. *Journal of Econometrics 43*, 371–387.

Bowman, A., M. Jones, and I. Gijbels (1998). Testing monotonicity of regression. *Journal of Computational and Graphical Statistics 7*(4), 489–500.

Chipman, J., D. McFadden, and M. Richter (Eds.) (1990). *Preferences, Uncertainty and Optimality*, Boulder. Westview Press.

Daly, A. and S. Zachary (1979). Improved multiple choice models. In D. Hensher and Q. Dalvi (Eds.), *Determinants of Travel Choice*, London, pp. 337–357. Teakfield.

Gourieroux, C., J. Laffont, and A. Montfort (1980). Coherency conditions in simultaneous linear equation models with endogenous switching regimes. *Econometrica 48*, 675–695.

Heckman, J. (1978). Dummy exogenous variables in a simultaneous equation system. *Econometrica 46*, 931–959.

Houthakker, H. (1950). Revealed preference and the utility function. *Economica 17*, 159–174.

Koning, R. and G. Ridder (1994). On the compatibility of nested logit models with utility maximization. *Journal of Econometrics 63*, 389–396.

Koning, R. and G. Ridder (2000). Utility maximization and mode of payment. In R. Heijmans, D. Pollock, and A. Satorra (Eds.), *Innovations in Multivariate Statistical Analysis*, Dordrecht, pp. 117–136. Kluwer Academic Publishers.

Luce, R. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: John Wiley & Sons.

Maddala, G. (1983). *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Manski, C. (1988a). *Analog Estimation Methods in Econometrics*. London: Chapman and Hall.

Manski, C. (1988b). Identification of binary response models. *Journal of the American Statistical Association 83*, 729–738.

Matzkin, R. (1992). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica 60*(2), 239–270.

Matzkin, R. (1993). Nonparametric identification and estimation of polychotomous chouce models. *Journal of Econometrics 58*, 137–168.

McFadden, D. (1978). Modelling the choice of residential location. In A. Karlquist (Ed.), *Spatial Interaction Theory and Residential Location*, Amsterdam, pp. 75–96. North-Holland.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 198–272. Cambridge: MIT Press.

McFadden, D. and M. Richter (1990). Stochastic rationality and revealed stochastic preference. See Chipman, McFadden, and Richter (1990), pp. 161–186.

Silverman, B. (1986). *Density Estimation*. London: Chapman & Hall.

Stoker, T. (1991). Equivalence of direct, indirect and slope estimators of average derivatives. In W. Barnett, J. Powell, and G. Tauchen (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge. Cambridge University Press.

Stoker, T. (1992). *Lectures on Semiparametric Econometrics*. Louvain-la-Neuve: CORE Foundation.

Thurstone, L. (1927). A law of comparative judgement. *Psychological Review 34*, 251–259.

Varian, H. (1984). *Microeconomic Analysis, Second Edition*. New York: W.W. Norton & Company.

Williams, H. (1977). On the formation of travel demand models and econometric evaluation measures. *Environment and Planning A 9*, 285–344.

Preference ordering	Probability	Event in ARUM
$1 \succ 2 \succ 3$	π_{123}	$D_{123} : w_1 < z_1, w_2 - w_1 < z_2 - z_1$
$1 \succ 3 \succ 2$	π_{132}	$D_{132} : w_2 < z_2, w_1 - w_2 < z_1 - z_2$
$2 \succ 1 \succ 3$	π_{213}	$D_{213} : w_1 > z_1, w_2 < z_2$
$2 \succ 3 \succ 1$	π_{231}	$D_{231} : w_2 - w_1 < z_2 - z_1, w_2 > z_2$
$3 \succ 1 \succ 2$	π_{312}	$D_{312} : w_2 > z_2, w_1 < z_1$
$3 \succ 2 \succ 1$	π_{321}	$D_{321} : w_1 > z_1, w_1 - w_2 < z_1 - z_2$

Table A-1. Preference orderings with probabilities and corresponding events in ARUM. $w_1 = \varepsilon_2 - \varepsilon_1, w_2 = \varepsilon_3 - \varepsilon_1, z_1 = v_2 - v_1, z_2 = v_3 - v_1$.

A. Proof of Theorem 1

Theorem 1. *Every ARUM with $\mathcal{V} = \{v\}$ implies a random preference model $(\mathcal{R}, \Pi)$. Conversely, every probability distribution Π over $\mathcal{R}$ can be represented by an ARUM with $\mathcal{V} = \{v\}$ for an appropriate choice of F.*

Proof. Consider an ARUM with non-random utility components v. Let R_k be an arbitrary complete and transitive strict preference ordering of the I alternatives. Hence, R_k can be written as

$$k_1 \prec k_2 \prec k_3 \prec \cdots \prec k_I$$

with $\{k_1, \ldots, k_I\}$ some permutation of $\{1, \ldots, I\}$. Define

$$\pi_k = \Pr(u_{k_1} < u_{k_2} < \cdots < u_{k_I}),\ k = 1, \ldots, I!.$$

Because the distribution of ε, and as a consequence, that of u, is absolutely continuous and non-defective, we have that

$$\Pr(u_i = u_j) = 0,\ i \neq j = 1, \ldots, I,$$

and, hence

$$\sum_{k=1}^{I!} \pi_k = 1.$$

Next, we prove the reverse assertion. If $I = 2$, the ARUM with non-random utility components $(v_1, v_2)'$ assigns $\Pr(\varepsilon_2 - \varepsilon_1 < v_2 - v_1) = \pi_{12}$ to the event that alternative 1 is strictly preferred over alternative 2. Hence, if we choose

$$f(\varepsilon_1, \varepsilon_2) = \begin{cases} \dfrac{\pi_{12}\phi(\varepsilon_2 - \varepsilon_1)\phi(\varepsilon_1)}{\Phi(v_2 - v_1)}, & \varepsilon_2 - \varepsilon_1 < v_2 - v_1 \\ \dfrac{\pi_{21}\phi(\varepsilon_2 - \varepsilon_1)\phi(\varepsilon_1)}{1 - \Phi(v_2 - v_1)}, & \varepsilon_2 - \varepsilon_1 > v_2 - v_1 \end{cases},$$

with π_{12} the probability that 1 is strictly preferred over 2, $\pi_{21} = 1 - \pi_{12}$, and ϕ the standard normal density function, then it is easily seen that the ARUM with this joint density function of $(\varepsilon_1, \varepsilon_2)'$ yields the probability distribution Π over the two strict preference orderings.

Next, we consider $I = 3$. The six possible preference orderings and associated probabilities are given in table A-1. An ARUM with the given non-random utility components v assigns utility levels

$$u_i = -v_i + \varepsilon_i,\ i = 1, 2, 3,y \tag{A-1}$$

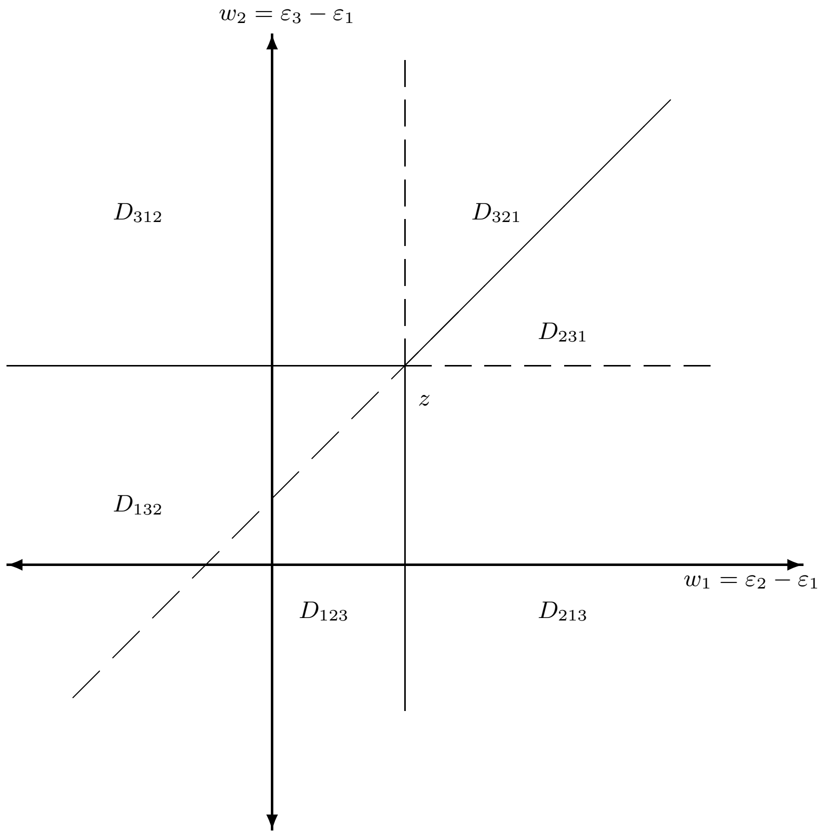

Figure A-1. Integration regions for preference orderings

to the three alternatives. According to the ARUM, the strict preference orderings in the first column of table A-1 obtain if and only if the events in the third column of the table occur. Hence, an ARUM that assigns the probabilities of the second column of table A-1 to the six strict preference orderings of column one is obtained, if we choose a joint density function of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)'$ such that the integrals over the indicated regions are equal to these probabilities. Note that we only have to find a joint density function of $(\varepsilon_2 - \varepsilon_1, \varepsilon_3 - \varepsilon_1)'$, because all events can be expressed in terms of these two random variables.

The integration regions in the third column of table A-1 are show in figure A-1. Let h, the joint density function of $(w_1, w_2)'$, be given by

$$h(w_1, w_2) = \frac{\pi_{ijk}\phi(w_1)\phi(w_2)}{\int\int_{D_{ijk}} \phi(s_1)\phi(s_2)\mathrm{d}s_1\mathrm{d}s_2},$$
$$(w_1, w_2) \in D_{ijk},\ i \neq j \neq k = 1, 2, 3,$$

with ϕ the standard normal density function. Then h is clearly a proper density function and we have:W

$$\int\int_{D_{ijk}} h(w_1, w_2)\mathrm{d}w_1\mathrm{d}w_2 = \pi_{ijk},\ i \neq j \neq k = 1, 2, 3.$$

It is obvious that if we choose the joint density function of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)'$

$$f(\varepsilon_1, \varepsilon_2, \varepsilon_3) = h(\varepsilon_2 - \varepsilon_1, \varepsilon_3 - \varepsilon_1)\phi(\varepsilon_1)$$

then the ARUM in equation (A-1) with this joint distribution yields the random preference model with probability distribution Π over the six strict preference orderings.

For $I \geq 4$, let R_k, $k = 1, \ldots, I!$ be all strict preference orderings of the I alternatives. Strict preference ordering R_k gives a complete ranking of the I alternatives

$$k_1 \prec k_2 \prec \cdots \prec k_I,$$

with $\{k_1, \ldots, k_I\}$ some permutation of $\{1, \ldots, I\}$. Define

$$\begin{aligned} D_k = \Big\{\varepsilon \in \mathbb{R}^I \mid (\varepsilon_{k_i} - \varepsilon_1) - (\varepsilon_{k_{i+1}} - \varepsilon_1) < (v_{k_i} - v_1) \\ -(v_{k_{i+1}} - v_1),\ i = 1, \ldots, I-1\Big\}\ k = 1, \ldots, I!.y \end{aligned} \tag{A-2}$$

Let

$$w = \varepsilon^1 - \iota_{I-1}\varepsilon_1,$$

with ε^1 the $(I-1)$-subvector of ε that does not contain the first component ε_1 and ι_{I-1} an $(I-1)$-vector of ones. Define

$$\tilde{D}_k = \left\{\varepsilon^1 - \iota_{I-1}\varepsilon_1 \in \mathbb{R}^{(I-1)} \mid \varepsilon \in D_k\right\},\ k = 1, \ldots, I!.$$

Note that

$$\bigcup_{k=1}^{I!} D_k = \mathbb{R}^I,\ D_k \bigcap D_l = \emptyset,\ k \neq l.$$

Hence, it is obvious that

$$\bigcup_{k=1}^{I!} \tilde{D}_k = \mathbb{R}^{(I-1)}.y \tag{A-3}$$

If $w \in \tilde{D}_k \bigcap \tilde{D}_l$, then there are $\varepsilon \in D_k$, $\tilde{\varepsilon} \in D_l$ with

$$\varepsilon^1 - \iota_{I-1}\varepsilon_1 = \tilde{\varepsilon}^1 - \iota_{I-1}\tilde{\varepsilon}_1.$$

From equation (A-2) we see that $\tilde{\varepsilon} \in D_k$, and we conclude that

$$\tilde{D}_k \bigcap \tilde{D}_l = \emptyset,\ k \neq l.y \tag{A-4}$$

The ARUM that assigns probabilities π_k to the strict preference ordering R_k, $k = 1, \ldots, I!$ has a joint density function of ε that can be constructed as follows. Define the density function

$$h(w) = \pi_k \frac{\prod_{i=1}^{I-1} \phi(w_i)}{\int_{D_k} \prod_{i=1}^{I-1} \phi(s_i)\mathrm{d}s},\ w \in \tilde{D}_k.$$

From equation (A-3) and equation (A-4) it follows that this is a proper density function. The joint density function of ε is:

$$f(\varepsilon) = h(\varepsilon_2 - \varepsilon_1, \ldots, \varepsilon_I - \varepsilon_1)\phi(\varepsilon_1)$$

B. Proof of Theorem 3

Proof. (Necessity) (C-1) follows directly from the uniqueness (with probability one) of the utility maximizing choice. Translation invariance is a direct consequence of equation (3). The non-defectiveness of the distribution of ε implies (C-3). The differentiability almost everywhere and non-negativity follow from equation (2) and the absolute continuity of the distribution of ε. Finally, we have

$$\frac{\partial P_j}{\partial v_i}(v) = \int_{-\infty}^{\infty} \frac{\partial^2 F}{\partial \varepsilon_i \partial \varepsilon_j}(\varepsilon_j - v_j + v_1, \ldots, \varepsilon_j, \ldots, \varepsilon_j - v_j + v_I)\mathrm{d}\varepsilon_j$$
$$= \int_{-\infty}^{\infty} \frac{\partial^2 F}{\partial \varepsilon_i \partial \varepsilon_j}(\varepsilon_i - v_i + v_1, \ldots, \varepsilon_i, \ldots, \varepsilon_i - v_i + v_I)\mathrm{d}\varepsilon_i$$
$$= \frac{\partial P_i}{\partial v_j}(v),$$

where the second equality is obtained by the change of variable $\varepsilon_i = \varepsilon_j + v_i - v_j$.
(Sufficiency) Translation invariance implies that we can write

$$P_i(v) = P_i(v - v_i\iota_I) = H_i(v^i - v_i\iota_{I-1})\mathrm{W} \tag{B-1}$$

with H_i a function defined on $\mathbb{R}^{(I-1)}$. Let for $w \in \mathbb{R}^{(I-1)}$

$$h_i(w) = \frac{\partial^{(I-1)} P_i}{(\partial w)^i}(w).y \tag{B-2}$$

Because of (C-4) and equation (B-1) h_i exists and is non-negative on $\mathbb{R}^{(I-1)}$. Moreover

$$P_i(v) = \int_{-\infty}^{v^i - v_i\iota_{I-1}} h_i(w)\mathrm{d}w.y \tag{B-3}$$

Note that from (C-5) for $i \neq j$ and all $v \in \mathbb{R}^I$

$$h_i(v^i - v_i\iota_{I-1}) = h_j(v^j - v_j\iota_{I-1}).y \tag{B-4}$$

Comparison of equation (B-3) and equation (3) indicates that we must show that there exists a random I-vector ε with an absolutely continuous and non-defective distribution such that for $i = 1, \ldots, I$

$$w^i = \varepsilon^i - \varepsilon_i\iota_{I-1}$$

has density function h_i.

Let k be an arbitrary non-defective density function and specify the distribution function of ε by

$$F(\varepsilon) = \int_{-\infty}^{\varepsilon_1} H_1(\varepsilon^1 - s\iota_{I-1})k(s)\mathrm{d}s.$$

The corresponding density function is, of course,

$$f(\varepsilon) = h_1(\varepsilon^1 - \varepsilon_1\iota_{I-1})k(\varepsilon_1).$$

Note that in the construction of F and f we started from 1 as a 'reference alternative'. A transformation of ε to w^1 and ε_1 shows that $\varepsilon^1 - \varepsilon_1\iota_{I-1}$ has density function h_1 (and the density function of ε_1 is k). We need to show that $\varepsilon^i - \varepsilon_i\iota_{I-1}$ has density function h_i for $i = 2, \ldots, I$.

Without loss of generality we choose $i = I$ (if necessary, we re-label the alternatives). Now consider the transformation from ε to

$$\begin{aligned} \eta_1 &= \varepsilon_1 - \varepsilon_I \\ &\vdots \\ \eta_{I-1} &= \varepsilon_{I-1} - \varepsilon_I \\ \eta_I &= \varepsilon_I \end{aligned} .$$

The corresponding density function of η is

$$g(\eta) = h_1(\eta_2 - \eta_1, \ldots, \eta_{I-1} - \eta_1, -\eta_1)k(\eta_1 + \eta_I). \tag{B-5}$$

It is easily seen that equation (B-4) implies that

$$h_1(\eta_2 - \eta_1, \ldots, \eta_{I-1} - \eta_1, \eta_I - \eta_1) = h_I(\eta_1 - \eta_I, \ldots, \eta_{I-1} - \eta_I).$$

Setting $\eta_I = 0$ and substituting in equation (B-5) gives

$$g(\eta) = h_I(\eta_1, \ldots, \eta_{I-1})k(\eta_1 + \eta_I).$$

Integrating out η_I shows that $\varepsilon^I - \varepsilon_I \iota_{I-1}$ has density function h_I.

The distribution of ε is absolutely continuous by construction (and the marginal distribution of ε_1 is by construction non-defective). It is also non-defective, because

$$F(\varepsilon) = \int_{-\infty}^{\varepsilon_1} P_1(0, \varepsilon^1 - s\iota_{I-1})k(s)\mathrm{d}s$$

and, hence from (C-3)

$$\lim_{\varepsilon^1 \to -\infty} F(\varepsilon) = 0,$$

$$\lim_{\varepsilon^1 \to \infty} F(\varepsilon) = \int_{-\infty}^{\varepsilon_1} k(s)\mathrm{d}s,$$

where the latter equality follows from

$$\lim_{v^1 \to \infty} P_1(0, v^1) = \lim_{v_1 \to -\infty} P_1(v - v_1\iota_I) = \lim_{v_1 \to -\infty} P_1(v) = 1.$$

C. A Comparison Between the Daly–Zachary Conditions and the Conditions for the Integrability of Demand Systems

The conditions (C-1), (C-4) and (C-5) resemble conditions that demand functions must satisfy to be compatible with utility maximization. It is well known that demand functions are compatible with utility maximization, if they have certain properties (see for instance Varian (1984)), as symmetry and non-negative definiteness of the Slutsky matrix. It is of interest to see how these properties compare to the Daly–Zachary conditions (C-1)–(C-5) above. For the sake of the analogy, we interpret v as the prices of the alternatives. We shall use some results of McFadden (1981) to derive a representative agent model that yields the choice probabilities as the demand functions in a continuous choice problem. We rewrite equation (1) as

$$u_{ti} = \frac{y_t}{p} - \frac{v_i}{p} + \varepsilon_{ti},\ i = 1, \ldots, I,\ t = 1, \ldots, T. \tag{C-1}$$

In equation (C-1) the subscript t refers to the t-th agent, y_t is his total expenditure and p is the price of other consumption expenditures. Note that adding $\frac{y_t}{p}$ does not affect the choice made by the agent. The only change in equation (1) is that we take the price of i relative to the price of other consumption. The representative agent has the following indirect utility and cost function:

$$\bar{V}(\bar{y}, v, p) = \frac{\bar{y}}{p} + E \max_{i=1,\dots,I} \left\{ -\frac{v_i}{p} + \varepsilon_i \right\},y \tag{C-2}$$

$$\bar{C}(u, v, p) = pu - E \max_{i=1,\dots,I} \{-v_i + p\varepsilon_i\},y \tag{C-3}$$

with $\bar{y}$ the arithmetic average of the total expenditures. The expectation is taken over the distribution of ε. Using similar arguments as McFadden (1981) we can show that equation (C-2) and equation (C-3) are a proper indirect utility and cost function that correspond to a choice problem in which $\bar{y}$ is divided over $I + 1$ goods with prices v and p. Using Roy's identity and Shephard's lemma we can derive the Marshallian and Hicksian demands which in this case coincide. It is not difficult to see that[11]

$$\frac{\partial \bar{C}}{\partial v}(u, v, p) = P\left(\frac{v}{p}\right),y \tag{C-4}$$

and the Marshallian demand for the other consumption is

$$\bar{x}(\bar{y}, v, p) = \frac{\bar{y} - P\left(\frac{v}{p}\right)' v}{p}.$$

In these expressions P is the I-vector of choice probabilities. The integrability conditions that the demand functions (C-4) must satisfy are

$$\frac{\partial P_i}{\partial v_j}(v) = \frac{\partial P_j}{\partial v_i}(v) \tag{C-5}$$

$$S = \left[\frac{\partial P_i}{\partial v_j}\right] \leq 0.y \tag{C-6}$$

If we compare these conditions with conditions (C-1)–(C-5) above, it is seen that translation invariance is not implied by equation (C-5) and equation (C-6). Moreover, equations (C-5) and (C-6) yield weaker restrictions on the choice probabilities than the symmetry condition (C-4) and the non-negativity condition (C-5) as the following example with the choice between two alternatives ($I = 2$) demonstrates. The Slutsky condition (C-6) requires that the matrix

$$\begin{pmatrix} \frac{\partial P_1}{\partial v_1} & \frac{\partial P_1}{\partial v_2} \\ \frac{\partial P_2}{\partial v_1} & \frac{\partial P_2}{\partial v_2} \end{pmatrix} \tag{C-7}$$

is negative semi-definite. In particular, the diagonal elements must be non-positive. The non-negativity condition (C-4) on the other hand requires that

$$\frac{\partial P_1}{\partial v_2} \geq 0 \text{ and } \frac{\partial P_2}{\partial v_1} \geq 0.$$

Since $P_1(v) + P_2(v) = 1$ and because of the symmetry condition (C-5) we have

$$\frac{\partial P_1}{\partial v_1} = -\frac{\partial P_2}{\partial v_1} = -\frac{\partial P_1}{\partial v_2} = \frac{\partial P_2}{\partial v_2} \leq 0.y \tag{C-8}$$

[11]Note that

$$\frac{\partial}{\partial v_j} \max_i \{-v_i + p\varepsilon_i\} = \begin{cases} -19 & j = \arg\max_i \{-v_i + p\varepsilon_i\} \\ 09 & j \neq \arg\max_i \{-v_i + p\varepsilon_i\} \end{cases}.$$

Hence, the Daly–Zachary–Williams conditions imply that the matrix in equation (C-7) is negative semi-definite and symmetric, that the off-diagonal elements are non-negative, and that the rows and columns of this matrix sum to 0. These conditions are stronger than the ones imposed by the integrability conditions (C-5) and (C-6). We conclude that compatibility with utility maximization by a representative agent yields weaker conditions on the choice probabilities than compatibility with individual utility maximization.

D. Proof of Theorem 5

We begin by introducing some additional notation. By a change of variables as in corollary 3 we see that, if choices are made by stochastic utility maximization, choice probabilities can be written as

$$P_i(v) = \int_{B_i(v)} h_1(w)\mathrm{d}w,$$

with

$$B_1(v) = \left\{w \in \mathbb{R}^{(I-1)} \mid w \leq v^1 - v_1\iota_{I-1}\right\}$$

$$B_i(v) = \left\{w \in \mathbb{R}^{(I-1)} \mid w_{i-1} \geq v_i - v_1, w_{j-1} - w_{i-1} \leq (v_j - v_1)+ \\ -(v_i - v_1),\ i \neq j = 2, \ldots, I\right\},\ i = 2, \ldots I.$$

In this formulation, alternative 1 is chosen as the reference alternative. Each observation v_t induces a partition of $\mathbb{R}^{(I-1)}$ into I disjoint sets $B_i(v_t)$:

$$\bigcup_{i=1}^{I} B_i(v_t) = \mathbb{R}^{(I-1)}$$

$$B_i(v_t) \cap B_j(v_t) = \emptyset,\ \ i \neq j.$$

For a given sample $v_t,\ t = 1, \ldots, T$, we define the sets C as the intersections:

$$C_{i_1 i_2 \cdots i_T} \equiv B_{i_1}(v_1) \cap B_{i_2}(v_2) \cap \ldots \cap B_{i_T}(v_T) \subset \mathbb{R}^{(I-1)} \tag{D-1}$$

for all $(i_1, i_2, \ldots, i_T)$ in the index set

$$J = \{(i_1, i_2, \ldots, i_T) \mid i_s = 1, \ldots, I,\ s = 1, \ldots, T\}.$$

The sets $C_{i_1 i_2 \cdots i_T}$ will be empty for many combinations of $i_1, i_2, \ldots, i_T$. For example, $B_1(v_1) \subset B_1(v_2)$ implies that $B_1(v_1) \cap B_i(v_2) = \emptyset$ for $i = 2, \ldots, I$. Furthermore, note that each set $B_i(v_t)$ can be written as the union of sets C:

$$B_i(v_t) = \bigcup_{J_i(v_t)} C_{i_1 i_2 \cdots i_T},$$

where the index set $J_i(v_t)$ is given by

$$J_i(v_t) = \{(i_1, i_2, \ldots, i_T) \mid i_t = i,\ i_s = 1, \ldots, I,\ s = 1, \ldots, T,\ s \neq t\}.$$

From now on, we restrict ourselves to those sets C which are not empty, *i.e.* those belonging to

$$\mathcal{C} \equiv \{C_{i_1 i_2 \cdots i_T} \mid C_{i_1 i_2 \cdots i_T} \neq \emptyset,\ i_t = 1, \ldots, I,\ t = 1, \ldots, T\}.$$

The corresponding index set is J^*, *i.e.*,

$$J^* = \{(i_1, i_2, \ldots, i_T) \in J \mid C_{i_1 i_2 \cdots i_T} \in \mathcal{C}\}.$$

The collection $\mathcal{C}$ is a partition of $\mathbb{R}^{(I-1)}$: the sets in $\mathcal{C}$ are disjoint and the union of all sets in $\mathcal{C}$ is $\mathbb{R}^{(I-1)}$. Using this, we can rewrite each observed choice probability as

$$P_i(v_t) = \int_{B_i(v_t)} h_1(w)\mathrm{d}w = \sum_{(i_1,i_2,\ldots,i_T)\in J_i^*(v_t)} \int_{C_{i_1 i_2 \cdots i_T}} h_1(w)\mathrm{d}w, y \tag{D-2}$$

where $J_i^*(v_t) = \{(i_1, i_2, \ldots, i_T) \in J_i(v_t) \mid C_{i_1 i_2 \cdots i_T} \in \mathcal{C}\}$. Finally we define

$$A_{i_1 i_2 \cdots i_T} = \int_{C_{i_1 i_2 \cdots i_T}} h_1(w)\mathrm{d}w. y \tag{D-3}$$

Now we are in a position to proof Theorem 5.

Proof. (Necessity) It is clear from equation (D-2) and equation (D-3) that all A's will be non-negative if a non-negative generating density function h_1 exists.
(Sufficiency) Suppose the set of equations (13) has a non-negative solution. It follows from equation (D-3) that we can construct a non-negative density $h_1(w)$ which generates the observed choice probabilities. Take for $(i_1, i_2, \ldots, i_T) \in J^*$ a non-negative function $g_{i_1 i_2 \cdots i_T}$, such that:

$$\int_{C_{i_1 i_2 \cdots i_T}} g_{i_1 i_2 \cdots i_T}(w)\mathrm{d}w = A_{i_1 i_2 \cdots i_T}.$$

One can choose

$$g_{i_1 i_2 \cdots i_T}(w) = A_{i_1 i_2 \cdots i_T} \frac{\prod_{j=1}^{I-1} \phi(w_j)}{\int\limits_{C_{i_1 i_2 \cdots i_T}} \prod_{j=1}^{I-1} \phi(w_j)\mathrm{d}w}, \qquad w \in C_{i_1 i_2 \cdots i_T}, y \tag{D-4}$$

with $\phi(\cdot)$ the standard normal density function. We define h_1^* by

$$h_1^*(w) = g_{i_1 i_2 \cdots i_T}(w),\ w \in C_{i_1 i_2 \cdots i_T},\ (i_1, i_2, \ldots, i_T) \in J^*.$$

It is clear that h_1^* is non-negative and that for $i = 1, \ldots, I,\ t = 1, \ldots, T$:

$$P_i(v_t) = \int_{B_i(v_t)} h_1^*(w)\mathrm{d}w.$$

If the density of ε is

$$h_1^*(\varepsilon_2 - \varepsilon_1, \ldots, \varepsilon_I - \varepsilon_1)\phi(\varepsilon_1)$$

then the choice probabilities can be written as in equation (9). From equation (D-4) it follows that we can choose the distribution of ε to be absolutely continuous and non-defective.

E. Compatibility on a Finite Set But Not on an Interval

In this section we apply the necessary and sufficient conditions of theorem 5 to choice probabilities that are generated by an NMNL model and we show that choice probabilities may be compatible on a finite set but not on an interval. To be specific the choice probabilities at the observed utility components v_t, $t = 1, \ldots, T$ are given by the nested multinomial logit (NMNL) model of McFadden (1978) (see also Maddala (1983), pp. 67-69). We consider an NMNL model with three alternatives ($I = 3$). The joint distribution of the random components of the utilities is

$$F(\varepsilon) = \exp\left\{-\left[\exp(-\varepsilon_1/\theta) + \exp(-\varepsilon_2/\theta)\right]^\theta - \exp(-\varepsilon_3)\right\}.$$

If we take the first alternative as the reference alternative, the distribution function of $w \equiv (\varepsilon_2 - \varepsilon_1, \varepsilon_3 - \varepsilon_1)'$ becomes

$$H_1(w) = \frac{(1 + \exp(-w_2/\theta))^{(\theta-1)}}{\exp(-w_1) + (1 + \exp(-w_2/\theta))^\theta} . y \tag{E-1}$$

The corresponding density function is

$$\begin{aligned} h_1(w_1, w_2) = {} & \exp(-w_2)\exp(-w_1/\theta)\frac{(1 + \exp(-w_1))^{(\theta-2)}}{\exp(-w_2) + (1 + \exp(-w_1/\theta))^\theta} \\ & \times \left\{\frac{2\,(1 + \exp(-w_1/\theta))^\theta}{\exp(-w_2) + (1 + \exp(-w_1/\theta))^\theta} - \frac{\theta - 1}{\theta}\right\}\frac{1}{\exp(-w_2) + (1 + \exp(-w_1/\theta))^\theta}. \end{aligned}$$

(*cf.* Börsch-Supan (1990), equations (13) and (14) which are not correct). This density is signed by the term in braces. The first term approaches 0 if $w_1 \to \infty$ and $w_2 \to -\infty$, so that the density is only nonnegative for for all $w \in \mathbb{R}^2$ if and only if $-\frac{\theta-1}{\theta} \geq 0$, *i.e.* if and only if $0 < \theta \leq 1$. In case θ is not in this interval, there is a set of positive measure where the function $h_1(w_1, w_2)$ is negative.

If $\theta > 1$, there exists a set of positive measure where $h_1(w)$ is negative. Hence, the choice probabilities do not satisfy the non-negativity condition in equation (C-4), and therefore they are not globally compatible with stochastic utility maximization. This is illustrated in figure E-1. There, and in the sequel we take $\theta = 2$. Now suppose we have a sample of three points ($T = 3$): $z_1 = (-1, 1)$, $z_2 = (2, 2)$ and $z_3 = (4, -2)$. The function $h_1(w)$ is negative in $z_3 = (4, -2)$.

Using equation (E-1), we can calculate the choice probabilities as

$$\begin{aligned} P(z_1) &= (0.36, 0.59, 0.05)' \\ P(z_2) &= (0.68, 0.25, 0.07)' \\ P(z_3) &= (0.13, 0.02, 0.85)'. \end{aligned}$$

After inspection of these choice probabilities (and figure E-1), it is seen that they satisfy the necessary condition of corollary 4. Moreover, using the notation of the preceding section, we see that the choice probabilities also satisfy the necessary and sufficient condition of theorem 5: a non-negative solution for the A's is $A_{111} = 0.13, A_{113} = 0.23, A_{211} = 0, A_{213} = 0.32, A_{221} = 0, A_{222} = 0.02, A_{223} = 0.23, A_{233} = 0.02, A_{313} = 0$ and $A_{333} = 0.05$. Hence we conclude that the observations are compatible with stochastic utility maximization, even though $h_1(w)$ is negative in z_3. Note that by theorem 4 the choice probabilities are *not* locally compatible on any interval that contains z_1, z_2, and z_3.

Now suppose we had another observation, say $z_4 = (3, -3)$. This observation has choice probabilities $(0.06, 0.01, 0.93)'$ according to the NMNL model. It is clear that (see figure E-1) $B_2(v_3) \subset B_2(v_4)$, but $P_2(v_3) > P_2(v_4)$, violating the necessary condition of corollary 4. This is, of course, due to the negativity of the density function in $B_2(v_4) \setminus B_2(v_3)$. It is no longer possible to find a density function $h_1(w)$ which could have generated the observed choice probabilities.

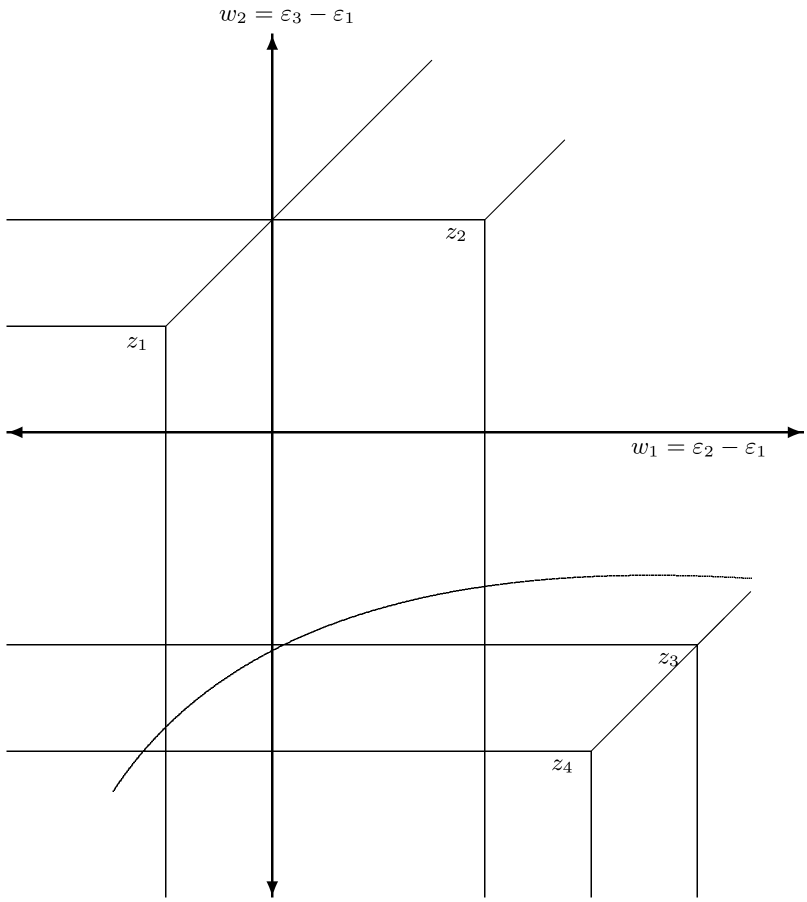

Figure E-1. Choice probabilities of the nested multinomial logit model