Chapter 4

Discrete Choice and Stochastic Utility Maximization

4.1 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 always can 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.

Let us observe the choice made by the agent. How 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 chapter, 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 $I$-ivariate distribution, of which we know the mean. Hence, we can write

$$ u_i = v_i + \varepsilon_i, \quad i = 1, \ldots, I $$

(4.1)
where $-v_i$ is the known mean of $u_i$, and $\varepsilon_i$ is a zero mean random variable. The joint distribution function of the $I$-vector $\varepsilon$ is denoted by $F$. For the moment we assume that this joint distribution function is known, possibly up to a vector of parameters. The model in equation (4.1) is the well-known (additive) random utility model.

Imperfect knowledge of the utilities of the alternatives makes it harder to predict the behaviour of the agent. In general, the choice made by the agent depends on the realization of the utilities $u_i, i = 1, \ldots, I$. If we assume that the agent has a decision rule based on the utilities, then 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.

By observing choices for a given vector of non-random utility components $v$ we can identify the choice probabilities $P_i(v), i = 1, \ldots, I$. If the random components $\varepsilon_i$ reflect intrapersonal variation in preferences, then we can identify the choice probabilities by observing repeated choices of a single agent. If the random components reflect interpersonal variation in preferences, then identification results from observing choices made by a group of agents with identical non-random utility components $v$. In econometric practice, we have estimates of the vectors $v_t, t = 1, \ldots, T$ for $T$ agents, and corresponding estimates of the choice probabilities $P_i(v_t), i = 1, \ldots, I, t = 1, \ldots, T$. Unbiased estimates of the choice probabilities can be obtained either by stratifying the sample with respect to $v_t$ or by fitting a flexible functional function, e.g. an unrestricted Nested Multinomial Logit model (see for instance chapter 5), to the $v_t$'s and the corresponding observed choices. In the latter case we can extrapolate the choice probabilities outside the set of observed $v_t$'s, to obtain estimates of the choice probabilities in some subset $\mathcal{V}$ of $\mathbb{R}^I$.

Although econometric practice is an important motivation for this chapter, we shall not discuss the details of estimation. We assume throughout that we have estimates of the choice probabilities $P_i(v), i = 1, \ldots, I$ for $v$ in some subset $\mathcal{V}$ of $\mathbb{R}^I$.

Given the choice probabilities for $v$ in some subset $\mathcal{V}$ of $\mathbb{R}^I$, can we test whether the agents that make these choices are rational, i.e. 1. Kernel smoothing can be seen as a combination of the two approaches.
that on all occasions they choose the alternative with the highest level of utility? In other words, we are interested in necessary and sufficient conditions for the compatibility of choice probabilities with maximization of the random utility function, i.e. with stochastic utility maximization. In this chapter we shall derive and discuss such conditions for various choices of $V$.

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 recent 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 make on all occasions a choice from all $I$ alternatives. However, the $I$-variate distribution of utility levels differs, either over time between choices of a single agent, or over agents between choices of distinct agents, due to differences in the non-random components $v$. Note that only if $v_i$ approaches infinity, alternative $i$ is eliminated from the choice set.

Our results follow up 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. A first attempt is reported in chapter 5.
The plan of the chapter is as follows. In section 4.2 we study the additive random utility model of equation (4.1). Section 4.3 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 4.4 we derive conditions for local compatibility for two classes of sets of non-random components \( V \). Section 4.5 contains some questions 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.

### 4.2 Additive Random Utility, Stochastic Preferences, and Choice Probabilities

In equation (4.1) 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 \( \varepsilon \) does not depend on \( \varphi \) for all \( \varphi \in V \).

This assumption is vacuous, if \( V \) contains only one element. The independence assumption is analogous to independence assumptions that are routinely made in regression models, including the latent regression models that are used to construct models for limited dependent variables. The independence assumption also allows us to recover \( F \) from the choice probabilities, i.e., given sufficient variation in \( \varphi \) the ARUM can be non-parametrically identified from the choice probabilities\(^2\). The independence assumption states that the utility of alternative \( i \) can be written as the sum of a systematic component \(-v_i\) and a purely random component \( \varepsilon_i \).

The other assumptions on the distribution of \( \varepsilon \) are

**Assumption 2. (absolute continuity.)** The joint distribution of \( \varepsilon \) is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^I \). In

\(^2\) Examples where the distribution function \( F \) is non- or semi-parametrically estimated are Cosslett (1983) and Klein and Spady (1993). Manski (1988) discusses nonparametric identification in the case where the choice set consists of only two alternatives.
other words, any (Borel-measurable) set of measure 0 according to the Lebesgue measure is assigned measure 0 under \( F \). Hence, we assume that \( F \) has no atoms.

**Assumption 3. (non-defectiveness.)** The joint distribution of \( \varepsilon \) is non-defective, i.e.,
\[
\lim_{\varepsilon \to -\infty} F(\varepsilon) = 1,
\]
\[
\lim_{\varepsilon \to +\infty} F(\varepsilon) = 0
\]
for \( i = 1, \ldots, I \).

A consequence of the assumptions of absolute continuity and non-defectiveness of \( F \) is that
\[
\Pr(u_i = u_j) = 0, \; i \neq j = 1, \ldots, I,
\]
i.e. the probability of ties is 0.

An ARUM is defined as a random utility model of the form
\[
\mu = -v + \varepsilon, \; v \in \mathbb{V}
\]
with \( u, v \), and \( \varepsilon \) \( I \)-vectors, and where the distribution function of \( \varepsilon \), \( F \), satisfies the independence, absolute continuity and non-defectiveness assumptions. Hence, an ARUM model is characterized by the triple \((I, F, \mathbb{V})\).

Now 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 \( \pi_i \) to all \( I! \) preference orderings in \( \mathcal{R} \), i.e. it consists of a pair \((\mathcal{R}, \Pi)\), with \( \Pi \) 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.

---

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 vacuous if \( \mathbb{V} \) is bounded.
Theorem 4.1. Every ARUM with \( V = \{v\} \) implies a random preference model \((R, \Pi)\). Conversely, every probability distribution \( \Pi \) over \( R \) can be represented by an ARUM with \( V = \{v\} \) for an appropriate choice of \( F \).

**Proof** See Appendix 4.A. \( \square \)

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 \( \varepsilon \) 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 to let \( \varepsilon \) represent interpersonal variation in preferences. Usually, the \( \varepsilon \) 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 chapter 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 \( \varepsilon \). 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 \( \varepsilon \). 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 perfect 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 \( \varepsilon \), 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

$$P_i(v) = \Pr(u_i > u_j, i \neq j = 1, \ldots, I) =$$

$$\int_{-\infty}^{\infty} \frac{\partial F}{\partial \varepsilon_i}(\varepsilon_i - v_i + v_j, \ldots, \varepsilon_i, \ldots, \varepsilon_i - v_i + v_j) d\varepsilon_i,$$

$$i = 1, \ldots, I. \quad (4.2)$$

The last equality follows from the absolute continuity of $F$. Note that the choice probabilities in equation (4.2) are invariant under a common strictly increasing transformation of the utility levels. Hence, the ARUM utility function is an ordinal utility function.

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 $\varepsilon$, the observable outcome is the chosen alternative and the mapping of $\varepsilon$ to $\{1, \ldots, I\}$ is the subscript of 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 $\varepsilon_1$ is arbitrary, we can choose it to be absolutely continuous

$^4$. This occurs if the distribution of $\varepsilon_2 - \varepsilon_1$ has an atom. It is equivalent to the inclusion of nonstrict preferences in $\mathcal{R}$. 
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 continuous⁵.

We have proved the following theorem.

**Theorem 4.2.** 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 $(\varepsilon_2 - \varepsilon_1 \ldots \varepsilon_t - \varepsilon_1)^\top$ is absolutely continuous with respect to the Lebesgue measure.

Next, we discuss the independence assumption. As usual, the routinely, and often implicitly, made independence assumption is hard to justify. It excludes e.g., some forms of heteroscedasticity of the random utility components. We have not found an explicit discussion of this assumption in the literature. McFadden (1981) uses characteristics of the agents to divide the population into homogeneous groups, i.e., groups with the same distribution of $\varepsilon$, and he identifies $v$ with costs associated with the alternatives, making the implicit assumption that costs vary independently of the characteristics of the agents and the alternatives. In general, it seems a good idea to impose some structure on $v$ and to separate systematic variation in the utilities due to e.g., variation in costs and characteristics of the alternatives from variation in the characteristics of the agents. Tests for rational choice can then be based on systematic variation in the utilities. By appropriate stratification on the characteristics of the agents we obtain in each stratum a distribution of $\varepsilon$ that does not depend on $v$. Of course, we require that there is sufficient variation in the utilities within the strata. This approach is in our opinion more fruitful than allowing for an ad hoc, and necessarily restrictive, dependence of $\varepsilon$ on $v$. As we shall see below, it is possible that the resulting choice probabilities lose appealing properties.

Whatever the justification, we do maintain the independence assumption. Hence, if choice probabilities do not satisfy our compatibility conditions, it is still possible that an (additive) random utility

---

⁵ Note, however, that distributions exist that yield a coherent stochastic utility maximization model even though they are not absolutely continuous. For example, let $\varepsilon_1$ have a discrete distribution and let $\varepsilon_2$ have a continuous distribution. The distribution of $\varepsilon_1 - \varepsilon_1$ will be continuous and the resulting stochastic utility maximization model will be coherent even though the joint distribution of $(\varepsilon_1, \varepsilon_2)^\top$ is not absolutely continuous.
model with $\varepsilon$ dependent on $v$ has generated the choice probabilities. However, we shall show that the assumption of stochastic utility maximization restricts the dependence of $\varepsilon$ on $v$ in additive models and that random parameter models that allow for heteroscedastic $\varepsilon$ do not satisfy this restriction if the heteroscedasticity depends on $v$.

Without loss of generality we set $I = 2$. The utilities of the two alternatives are

$$u_1 = -v_1 + \varepsilon_1, \quad u_2 = -v_2 + \varepsilon_2, \quad v \in V, \quad (4.3)$$

where the joint distribution of $\varepsilon$ may depend on $v$. The absolute continuity and non-defectiveness assumptions are maintained. The choice probability of alternative 1 is

$$P_1(v) = \Pr(\varepsilon_2 - \varepsilon_1 < v_2 - v_1) = H(v_2 - v_1 \mid v), \quad v \in V \quad (4.4)$$

with $H$ the distribution function of $\varepsilon_2 - \varepsilon_1$ which may depend on $v$. From equation (4.3) it is obvious that the choice of a rational agent is not affected by the addition of a common constant to the non-random utility components. Thus for all $c \in \mathbb{R}$

$$P_1(v_1 + c, v_2 + c) = P_1(v_1, v_2), \quad i = 1, 2$$

i.e. the choice probabilities are translation invariant. Note that the choice probabilities are translation invariant if and only if they are a function of $v_2 - v_1$ only. It follows from equation (4.4) that this holds if and only if the distribution of $\varepsilon_2 - \varepsilon_1$ depends on $v$ through $v_2 - v_1$. A similar result holds if $I \geq 3$.

Not all random utility models are translation invariant. Let us consider a simple random parameter model that allows for heteroscedastic random utility components. The utility function is

$$u_i = \beta_i x_i, \quad \beta_i = \beta + \eta_i, \quad i = 1, \ldots, I \quad (4.5)$$

where $\eta$ has a joint distribution function $G$ that does not depend on $x$, and $\beta$ and $x_i$ are non-stochastic. We can write equation (4.5) in the ARUM form

$$u_i = \beta x_i + \eta_i x_i = -v_i + \varepsilon_i, \quad i = 1, \ldots, I.$$
\[ P_i(v) = \int_{-\infty}^{\infty} \frac{\partial G}{\partial \eta_i} \left( -\beta + (\beta + \eta_i) \frac{v_i}{v_1}, \ldots, \eta_i, \ldots, -\beta + (\beta + \eta_i) \frac{v_i}{v_I} \right) d\eta_i \]

which is clearly not translation invariant.

### 4.3 Global Compatibility with Stochastic Utility Maximization

#### 4.3.1 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 section 4.4 of the present chapter.

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

**Definition 4.1.** 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

\[ P_i(v) = \Pr(\varepsilon_j - v_j < \varepsilon_i - v_i; i \neq j = 1, \ldots, I), \ i = 1, \ldots, I \quad (4.6) \]

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

In the sequel \( t_{I-1} \) denotes an \((I-1)\)-vector of ones. The \((I-1)\)-vectors \( \varepsilon^i \) and \( \varepsilon^j \) are obtained from the \( I \)-vectors \( \varepsilon \) and \( \varepsilon \) 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 \( v \in \mathbb{R}^I \):

\[ P_i(v) \geq 0, \quad \sum_{i=1}^{I} P_i(v) = 1 \quad (C-1) \]
Global Compatibility  

\[ P_i(v) = P_i(v + ct) \text{ for all } c \in \mathbb{R} \text{ (translation invariance)} \quad (C-2) \]

\[ \lim_{v_i \to -\infty} P_i(v) = 1, \quad \lim_{v_j \to -\infty} P_j(v) = 0 \quad (C-3) \]

\[ P_i(v) \text{ is differentiable with respect to } v^i \text{ and } \frac{\partial (v^{i-1})}{(\partial v)^i} P_i(v) \geq 0 \]

\[ \text{(non-negativity)} \quad (C-4) \]

\[ \frac{\partial P_i}{\partial v^i_j}(v) = \frac{\partial P_i}{\partial v^i_i}(v) \text{ (symmetry).} \quad (C-5) \]

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

**Theorem 4.3.** (Daly and Zachary) Conditions (C-1)–(C-5) are necessary and sufficient for global compatibility with stochastic utility maximization.

Before proving this theorem, we give an interpretation of these conditions. The first condition states that all choice probabilities are non-negative and that some alternative must be chosen. According to (C-2), only the differences in average utilities determine the choice probabilities, not the absolute levels. This does not imply that utility is cardinal. If the utilities \( u_i \) are transformed by some monotonic, increasing function, the same choice probabilities are obtained. 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.

**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 (4.6). The non-defectiveness of the distribution of \( \varepsilon \) implies (C-3). The differentiability almost everywhere and non-negativity follow from equation (4.2) and the absolute continuity of the distribution of \( \varepsilon \). Finally, we have

\[ \frac{\partial P_i}{\partial v^i_j}(v) = \int_{-\infty}^{\infty} \frac{\partial^2 F}{\partial \varepsilon_i \partial \varepsilon_j}(\varepsilon - v_j + v_1, \ldots, \varepsilon_j, \ldots, \varepsilon_j, v_i + v) \, d\varepsilon_j \]
\[ \begin{align*}
&= \int_{-\infty}^{\infty} \frac{\partial^2 F}{\partial \varepsilon_i \partial \varepsilon_j} (\varepsilon_i - v_i + v_1, \ldots, \varepsilon_i - v_i + v_I) d\varepsilon_i \\
&= \frac{\partial P_i}{\partial \varepsilon_j}(v),
\end{align*} \]

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

\[ P_i(v) = P_i(v - v_{ij}) = H_i(v^i - v_{ij}^{t_{t-1}}) \] (4.7)

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

\[ h_i(w) = \frac{\partial (I-1) P_i}{\partial (w)}(w). \] (4.8)

Because of (C-4) and equation (4.7) \( h_i \) exists and is non-negative on \( \mathbb{R}^{(I-1)} \). Moreover

\[ P_i(v) = \int_{-\infty}^{w^i - v_{ij}^{t_{t-1}}} h_i(w) dw. \] (4.9)

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

\[ h_i(v^i - v_{ij}^{t_{t-1}}) = h_j(v^j - v_{ij}^{t_{t-1}}). \] (4.10)

Comparison of equation (4.9) and equation (4.6) indicates that we must show that there exists a random \( I \)-vector \( \varepsilon \) with an absolutely continuous and non-defective distribution such that for \( i = 1, \ldots, I \)

\[ w^i = \varepsilon^i - \varepsilon_{i}^{t_{t-1}} \]

has density function \( h_i \).

Let \( k \) be an arbitrary non-defective density function and specify the distribution function of \( \varepsilon \) by

\[ F(\varepsilon) = \int_{-\infty}^{\varepsilon_1} H_i(\varepsilon^1 - s_{t_{t-1}}) k(s) ds. \]

The corresponding density function is, of course,

\[ f(\varepsilon) = h_i(\varepsilon^1 - \varepsilon_{i}^{t_{t-1}}) k(\varepsilon^1). \]

Note that in the construction of \( F \) and \( f \) we started from 1 as a ‘reference alternative’. A transformation of \( \varepsilon \) to \( w^1 \) and \( \varepsilon^1 \) shows that \( \varepsilon^1 - \varepsilon_{1}^{t_{t-1}} \)
has density function \( h_1 \) (and the density function of \( \varepsilon_1 \) is \( k \)). We need to show that \( \varepsilon^I - \varepsilon_i t_{t-1} \) has density function \( h_i \) for \( i = 2, \ldots, I \). Without loss of generality we choose \( i = I \) (if necessary, we relabel the alternatives).

Now consider the transformation from \( \varepsilon \) to

\[
\eta_i = \varepsilon_1 - \varepsilon_i
\]

\[
\vdots
\]

\[
\eta_{t-1} = \varepsilon_{t-1} - \varepsilon_i
\]

\[
\eta_t = \varepsilon_i
\]

The corresponding density function of \( \eta \) is

\[
g(\eta) = h_1(\eta_2 - \eta_1, \ldots, \eta_{t-1} - \eta_1, -\eta_t)k(\eta_t + \eta_t).
\tag{4.11}
\]

It is easily seen that equation (4.10) implies that

\[
h_1(\eta_2 - \eta_1, \ldots, \eta_{t-1} - \eta_1, \eta_t - \eta_1) = h_I(\eta_t - \eta_t, \ldots, \eta_{t-1} - \eta_t).
\]

Setting \( \eta_t = 0 \) and substituting in equation (4.11) gives

\[
g(\eta) = h_I(\eta_t, \ldots, \eta_{t-1})k(\eta_t + \eta_t).
\]

Integrating out \( \eta_t \) shows that \( \varepsilon^I - \varepsilon_i t_{t-1} \) has density function \( h_I \).

The distribution of \( \varepsilon \) is absolutely continuous by construction (and the marginal distribution of \( \varepsilon_1 \) is by construction non-defective). It is also non-defective, because

\[
F(\varepsilon) = \int_{-\infty}^{\varepsilon_1} P_1(0, \varepsilon^I - st_{t-1})k(s)ds
\]

and, hence from (C-3)

\[
\lim_{\varepsilon^I \to -\infty} F(\varepsilon) = 0,
\]

\[
\lim_{\varepsilon^I \to -\infty} F(\varepsilon) = \int_{-\infty}^{\varepsilon_1} k(s)ds,
\]

where the latter equality follows from

\[
\lim_{\varepsilon^I \to -\infty} P_1(0, \varepsilon^I) = \lim_{\varepsilon^I \to -\infty} P_1(\varepsilon - \varepsilon_i t_{t-1}) = \lim_{\varepsilon^I \to -\infty} P_1(\varepsilon) = 1.
\]

\[\square\]
Remark 4.1. The definition of global compatibility implies that there exists a stochastic $I$-vector $\varepsilon$ that satisfies equation (4.6). The proof shows that the choice of $\varepsilon$ is not unique. The choice probabilities determine $h_1$, and by equation (4.10) also $h_2, \ldots, h_I$. In other words, they determine the distributions of $\varepsilon^1 - \varepsilon^I_{t-1}, \ldots, \varepsilon^I - \varepsilon^I_{t-1}$. One marginal distribution, e.g., the distribution of $\varepsilon_1$, can be chosen arbitrarily. From equation (4.10) 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 4.2. 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 $\varepsilon$, i.e., using alternative 1 as a reference alternative (see (5.131), p. 263). This establishes directly that

$$P_i(v) = \Pr(\varepsilon^1 - \varepsilon^I_{t-1} < v^1 - v^I_{t-1}).$$

(4.12)

Next, he proves that equation (4.12) 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 (4.2). 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 4.3. Using theorem 4.3 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

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

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 4.4.** 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 (see the example in section 4.5).

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

$$P_i(v) \text{ is differentiable with respect to } v^i \text{ and } \frac{\partial^k P_i(v)}{\partial v^i t_{l-1}} \geq 0(C'4)$$

Hence, we have

**Corollary 4.1.** 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 4.3 contains a further useful corollary.

**Corollary 4.2.** 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

$$h_i(v^i - v^i t_{l-1}) = h_j(v^i - v^i t_{l-1}), \quad i \neq j \quad (4.13)$$

and

$$P_i(v) = \int_{-\infty}^{v^i - v^i t_{l-1}} h_i(w)dw. \quad (4.14)$$

By a change of variables we obtain a third corollary.

**Corollary 4.3.** 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

$$P_i(v) = \int_{-\infty}^{v^i - v^i t_{l-1}} h_1(w)dw, \quad (4.15)$$
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 \((\varepsilon_2 - \varepsilon_1, \ldots, \varepsilon_T - \varepsilon_1)^t\), where the first alternative is arbitrarily taken as a reference alternative.

### 4.3.2 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, only if they have certain properties (see for instance Varian (1984)), as the symmetry condition 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 (4.1) as

\[
... = \frac{y_t}{p} - \frac{v_i}{p} + \varepsilon_{ui}, \quad i = 1, \ldots, I, \quad t = 1, \ldots, T. \tag{4.17}
\]

In equation (4.17) 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 (4.1) 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:

\[
\bar{V}(\bar{y}, v, p) = \frac{\bar{y}}{p} + E \max_{i=1, \ldots, T} \left\{ -\frac{v_i}{p} + \varepsilon_i \right\}, \tag{4.18}
\]

\[
\bar{C}(u, v, p) = pu - E \max_{i=1, \ldots, T} \left\{ -v_i + p\varepsilon_i \right\}, \tag{4.19}
\]
with \( \bar{y} \) the arithmetic average of the total expenditures. Of course, the expectation is taken over the distribution of \( \varepsilon \). Using similar arguments as McFadden (1981) we can show that equation (4.18) and equation (4.19) are a proper indirect utility and cost function that correspond to a choice problem in which \( y \) 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
\[
\frac{\partial \tilde{C}}{\partial v}(u, v, p) = P \left( \frac{v}{p} \right),
\]
and the Marshallian demand for the other consumption is
\[
\bar{x}(\bar{y}, v, p) = \frac{\bar{y} - P \left( \frac{v}{p} \right)^	op v}{p}.
\]
In these expressions \( P \) is the \( I \)-vector of choice probabilities. The integrability conditions that the demand functions (4.20) must satisfy are
\[
\frac{\partial P_i}{\partial v_j}(v) = \frac{\partial P_i}{\partial v_i}(v) \quad (4.21)
\]
\[
S = \left[ \frac{\partial P_i}{\partial v_j} \right] \leq 0. \quad (4.22)
\]
If we compare these conditions with conditions (C-1)–(C-5) above, it is seen that translation invariance is not implied by equation (4.21) and equation (4.22). Moreover, equations (4.21) and (4.22) 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 (4.22) requires that the matrix
\[
\begin{pmatrix}
\frac{\partial P_1}{\partial v_1} & \frac{\partial P_1}{\partial v_2} \\
\frac{\partial P_2}{\partial v_1} & \frac{\partial P_2}{\partial v_2}
\end{pmatrix}
\]
(4.23)

Note that
\[
\frac{\partial}{\partial v_j} \max \{ -v_j + px \} = \begin{cases} -1 & j = \arg \max_i \{ -v_i + px \} \\ 0 & j \neq \arg \max_i \{ -v_i + px \} \end{cases}.
\]
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

\[ \frac{\partial P_1}{\partial v_1} = -\frac{\partial P_2}{\partial v_1} = -\frac{\partial P_1}{\partial v_2} = \frac{\partial P_2}{\partial v_2} \leq 0. \] (4.24)

Hence, the Daly-Zachary-Williams conditions imply that the matrix in equation (4.23) 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 (4.21) and (4.22). We conclude that compatibility with utility maximization by a representative agent yields weaker conditions on the choice probabilities than compatibility with individual utility maximization.

### 4.4 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 4.2.** The set of choice probabilities \( P_i(v), i = 1, \ldots, I \) is locally compatible with stochastic utility maximization on a set \( V \subset \mathbb{R}^I \), if for \( i = 1, \ldots, I \) and all \( v \in V \) we can write

\[ P_i(v) = \Pr(\varepsilon_j - v_j < \varepsilon_i - v_i; j = 1, \ldots, I, j \neq i) \] (4.25)

with \( \varepsilon \) a stochastic \( I \)-vector with a non-defective and absolutely continuous distribution that does not depend on \( v \).

Local compatibility was introduced by 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 $V$ with $v_t \in V$ 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 $a, b \in \mathbb{R}^J$ 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 $V = [a, b]^J$. The following theorem gives necessary and sufficient conditions.

**Theorem 4.4.** The choice probabilities $P_i(v)$, $i = 1, \ldots, I$ are locally compatible with stochastic utility maximization on a bounded interval $V = [a, b]$ if and only if for all $v \in [a, b]$ (C-1), (C-2), (C’-4) and (C-5) hold.

**Proof** See Appendix 4.B. □

**Remark 4.5.** This result is stronger than that obtained by Börsch-Supan (1990). First, we do not require that (C-1), (C-2) and (C-5) are globally satisfied. Second, the theorem gives necessary and sufficient conditions. Third, Börsch-Supan does not prove local compatibility as defined above. His suggested distribution of $(\varepsilon_2 - \varepsilon_1 \ldots \varepsilon_J - \varepsilon_1)'$ is not absolutely continuous with respect to the Lebesgue measure, and he does not indicate how the resulting ties will be resolved. The construction in appendix B yields an absolutely continuous distribution of $\varepsilon$. The proof uses the representation of the choice probabilities of corollary 4.3 of theorem 4.3. We find a density function $h^*_1$ that satisfies this equation for all $v \in [a, b]$.

**Remark 4.6.** Condition (C-3) is not required for local compatibility. Since equation (4.25) only has to be satisfied on an interval, we have

7. Strictly, local compatibility on a bounded interval is not possible: if equation (4.25) holds for all $v \in [a, b]$ it holds for all $v$ such that $v + \varepsilon_{1:t} \in [a, b]$ for some $\varepsilon \in \mathbb{R}$. Hence, we must choose a normalization. For that purpose, we express $v$ in deviation from $v_1$, and we take $V$ as the set $\{v \in \mathbb{R}^J | v - v_{1:t} \in [a, b]\}$. Because $a_1 = b_1 = 0$, it suffices to specify $V = [a^1, b^1] \subset \mathbb{R}^{J-1}$.
more freedom in choosing the distribution of $\epsilon$. In particular, we can always choose it to be non-defective.

**Remark 4.7.** The conditions in theorem 4.4 and corollary 4.1 of theorem 4.3 are identical, except for (C-3).

Theorem 4.4 gives necessary and sufficient conditions for local compatibility on an interval. It is natural to ask whether these conditions are necessary and sufficient for local compatibility on arbitrary sets $\mathcal{V}$. The following simple example shows that this is not true.

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. The choice probabilities $P_1(v)$ and $P_2(v)$ can be expressed as

$$P_1(v) = \int_{-\infty}^{v_2 - w_1} h_1(w)dw,$$

$$P_2(v) = \int_{v_2 - w_1}^{\infty} h_1(w)dw,$$

with $h_1(w)$ defined in equation (4.8). Let $h_1$ and $H_1$, defined in equation (4.7), be as in figure 4.1.

Note that $h_1$ is non-negative on $\bar{V}_1 = (-\infty, w_{1}] \cup [w_1, \infty)$. However, it is not possible to find a density function $h_1^*$ that coincides with $h_1$ on $\bar{V}_1$ and also satisfies

$$P_1(v) = \int_{-\infty}^{v_2 - w_1} h_1^*(w)dw,$$

$$P_2(v) = \int_{v_2 - w_1}^{\infty} h_1^*(w)dw,$$

The obvious problem is that

$$P_1(0, w_1) < P_1(0, w_{11})$$

and this implies that $h_1^*$ has to be negative for some values of $w$. We conclude that the non-negativity condition in (C'-4) is not sufficient for local compatibility on more general sets than intervals.

---

8. These assumptions apply if we fit a flexible functional form, that satisfies all conditions, except the non-negativity condition.
In this simple example it is not difficult to find sets $\hat{V}$ on which the choice probabilities are locally compatible. From figure 4.1 we see that local compatibility holds on either $\hat{V}_2 = (-\infty, w_3] \cup [w_1, \infty)$ or $\hat{V}_3 = (-\infty, w_3] \cup [w_2, \infty)$. Actually, local compatibility holds on sets $\hat{V}_4 = (-\infty, w^*] \cup [w''', \infty)$ if and only if

$$P_1(0, w') \leq P_1(0, w'').$$

This example shows the limitations of theorem 4.4 in checking local compatibility. Assume that we have observed utilities $v_t$, $t = 1, \ldots, T$. If $v_{2t} - v_{1t} \in \hat{V}_1$ for $t = 1, \ldots, T$ we would conclude from theorem 4.4 that the choice probabilities are compatible with stochastic utility maximization. Of course, this conclusion is incorrect. If $v_{2t} - v_{1t} \in \hat{V}_2$ or $v_{2t} - v_{1t} \in \hat{V}_3$ for all $t$, we correctly conclude that the choice probabilities are locally compatible. Note that theorem 4.4 in this case does not lead to the conclusion that the choice probabilities are compatible with stochastic utility maximization, because the interval containing $\hat{V}_2$ or $\hat{V}_3$ is $(-\infty, \infty)$, and $h_1$ is negative on $(w_0, w_1)$. From the choice of $\hat{V}_4$ we see that even if $v_{2t} - v_{1t} \notin \hat{V}_1$, i.e. if $h_1(v_{2t} - v_{1t}) < 0$, for some $t$, there still

---

9. Because we express $v$ as deviation from $v_1$, it suffices to specify $a_2$ and $b_2$ ($a_2 = b_1 = 0$).
is hope that the choice probabilities are locally compatible. A necessary and sufficient condition in the case $I = 2$ is that for all $s, t = 1, \ldots, T$ we have

$$v_{2t} - v_{1t} \geq v_{2s} - v_{1s} \Rightarrow P_1(v_{1t}, v_{2t}) \geq P_1(v_{1t}, v_{2s}).$$  \hspace{1cm} (4.26)

Note that condition (4.26) does not preclude that $w_0 \leq v_{2t} - v_{1t} \leq w_1$ for some $t$. In other words, theorem 4.4 is also not necessary for local compatibility of the observed choice probabilities. We conclude that if one has observed utilities $v_t, t = 1, \ldots, T$, it is restrictive to investigate local compatibility by finding $a_2$ and $b_2$ such that $a_2 \leq v_{2t} - v_{1t} \leq b_2$ and checking whether theorem 4.4 holds for $V = [a_2, b_2]$.

We conclude that theorem 4.4 can only be applied if $V$ is a (bounded) interval. So is it possible to find similar necessary and sufficient conditions for other choices of $V$? We shall consider a choice of $V$ that can be seen as the opposite extreme, namely a finite set of distinct points. This choice of $V$ is of considerable practical interest, because in practice an econometrician has a finite sample $V = \{v_1, \ldots, v_T\}$ of observed utility components. 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 MNM 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 compatible with stochastic utility maximization, i.e. how can he decide whether the choice probabilities are locally compatible with stochastic utility maximization on $V$? In theorem 4.5 we give necessary and sufficient conditions for local compatibility on a set that consists of a finite number of distinct points in $\mathbb{R}^I$.

The derivation of these conditions is facilitated by some additional notation. By a change of variables as in corollary 4.3 we see that, if choices are made by stochastic utility maximization, choice probabilities can be written as

$$P_i(v) = \int_{B_i(v)} h_i(w)dw,$$

with

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

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 \( B_i(v_t) \):

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

\[ B_i(v_t) \cap B_j(v_t) = \emptyset, \quad i \neq j. \]

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

\[ C_{i_1, i_2, \ldots, i_T} \equiv B_{i_1}(v_1) \cap B_{i_2}(v_2) \cap \ldots \cap B_{i_T}(v_T) \subset \mathbb{R}^{(I-1)} \quad (4.27) \]

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

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

This notation is clarified in the example given in figure 4.2. There we have for example \( C_{213} = B_2(v_1) \cap B_1(v_2) \cap B_2(v_3) \). The sets \( C_{i_1, i_2, \ldots, i_T} \) 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 \( B_1(v_1) \cap B_1(v_2) = \emptyset \) for \( i = 2, \ldots, I \). Furthermore, note that each set \( B_i(v_t) \) can be written as the union of sets \( C \):

\[ B_i(v_t) = \bigcup_{J_i(v_t)} C_{i_1, i_2, \ldots, i_T}, \]

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

\[ C \equiv \left\{ C_{i_1, i_2, \ldots, i_T} \mid C_{i_1, i_2, \ldots, i_T} \neq \emptyset, i_t = 1, \ldots, I, t = 1, \ldots, T \right\}. \]

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

$$P_i(v_t) = \int_{B_i(v_t)} h_i(w)dw = \sum_{(i_1, i_2, \ldots, i_{I-1}) \in J^*_i(v_t)} \int_{C_{i_1,i_2,\ldots,i_{I-1}}} h_1(w)dw,$$

(4.28)

where $J^*_i(v_t) = \{(i_1, i_2, \ldots, i_{I-1}) \mid (i_1, i_2, \ldots, i_{I-1}) \in J_i(v_t) \cap C_{i_1,i_2,\ldots,i_{I-1}} \in C\}$. As a last bit of notation we define

$$A_{i_1,i_2,\ldots,i_{I-1}} = \int_{C_{i_1,i_2,\ldots,i_{I-1}}} h_1(w)dw.$$  

(4.29)

Now we are in a position to give a necessary and sufficient condition for local compatibility of the choice probabilities on $V$:

**Theorem 4.5.** The choice probabilities $P_i(v), i = 1, \ldots, I$ are locally compatible with stochastic utility maximization on $V = \{v_1, \ldots, v_T\}$ if and only if condition (C-1) holds on $V$ and the system of equations in $A_{i_1,i_2,\ldots,i_{I-1}}$

$$P_i(v_t) = \sum_{(i_1, i_2, \ldots, i_{I-1}) \in J^*_i(v_t)} A_{i_1,i_2,\ldots,i_{I-1}}$$

(4.30)

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

**Proof** (Necessity) It is clear from equation (4.28) and equation (4.29) that all $A_i$'s will be non-negative if a non-negative generating density function $h_1$ exists.

(Sufficiency) Suppose the set of equations (4.30) has a non-negative solution. It follows from equation (4.29) that we can construct a non-negative density $h_1(w)$ which generates the observed choice probabilities. Take for $(i_1, i_2, \ldots, i_{I-1}) \in J^*$ a non-negative function $g_{i_1,i_2,\ldots,i_{I-1}}$, such that:

$$\int_{C_{i_1,i_2,\ldots,i_{I-1}}} g_{i_1,i_2,\ldots,i_{I-1}}(w)dw = A_{i_1,i_2,\ldots,i_{I-1}}.$$

One can choose

$$g_{i_1,i_2,\ldots,i_{I-1}}(w) = A_{i_1,i_2,\ldots,i_{I-1}} \frac{\prod_{j=1}^{I-1} \phi(w_j)}{\int_{C_{i_1,i_2,\ldots,i_{I-1}}} \prod_{j=1}^{I-1} \phi(w_j)dw}, \quad w \in C_{i_1,i_2,\ldots,i_{I-1}},$$

where $\phi(w)$ is a non-negative function with $\int_{C_{i_1,i_2,\ldots,i_{I-1}}} \phi(w)dw = 1$.
with \( \phi(\cdot) \) the standard normal density function. We define \( h_1^* \) by

\[
h_1^*(w) = g_{i_1, i_2, \ldots, i_T}(w), \quad w \in C_1, i_2, \ldots, i_T) \in J^*.
\]

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

\[
P_i(v_t) = \int_{B_t(v_t)} h_1^*(w) \, dw.
\]

If the density of \( \varepsilon \) is

\[
h_1^*(\varepsilon_2 - \varepsilon_1, \ldots, \varepsilon_T - \varepsilon_1) \phi(\varepsilon_1)
\]

then the choice probabilities can be written as in equation (4.25). From equation (4.31) it follows that we can choose the distribution of \( \varepsilon \) to be absolutely continuous and non-defective.

**Remark 4.8.** The existence of a solution to equation (4.30) 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 (4.30) has a non-negative solution.

**Remark 4.9.** The equation system in equation (4.30) 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 4.5 is different from the McFadden/Richter test. The McFadden/Richter test exploits variation in the choice probabilities if the choice set is restricted, while in the test of theorem 4.5 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 trials \( (B, C_i) \) 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 4.5, we give a necessary condition on the choice probabilities that can be easily checked.

**Corollary 4.4.** If the choice probabilities \( P_i(v) \) are compatible with stochastic utility maximization, then for each pair \((t, t')\):
\[ B_i(v_i) \subseteq B_i(v_{i'}) \Rightarrow P_i(v_i) \leq P_i(v_{i'}). \] (4.32)

If \( I = 3 \), it is easily seen that each pair of points \((t, t')\) yields two restrictions of the form (4.32) (see also figure 4.2). Hence with \( T \) observations there will be \( T \times (T-1) \) restrictions like (4.32).

If condition (4.32) is violated for some pair \((t, t')\), one can conclude that the choice probabilities are not locally compatible with stochastic utility maximization. If \( I = 2 \), the condition of corollary 4.4 guarantees that the distribution function of \( \varepsilon_2 - \varepsilon_1 \) is non-decreasing in \( v_{2t} - v_{1t} \), \( t = 1, \ldots, T \) as, of course, it should be. For \( I = 2 \) the condition of corollary 4.4 is also sufficient for local compatibility with stochastic utility maximization. This result, however, does not generalize to more alternatives, as the following example illustrates\(^\text{10}\).

**Example** Condition (4.32) is not sufficient. Consider the case \( I = 3 \), \( T = 3 \). Let the first alternative be the reference alternative, and let \( v_i' = v_{1t} \). The points \( \tilde{w}_1, \tilde{w}_2, \tilde{w}_3 \) and the integration region of the corresponding choice probabilities are given in figure 4.2. The integration regions are decomposed in sets \( C_{i,i't,i't} \) defined in equation (4.27). Let the corresponding 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 4.4:

\[
\begin{align*}
B_1(v_1) & \subseteq B_i(v_2) \Rightarrow P_1(v_1) \leq P_1(v_2), \ (0 \leq 0) \\
B_2(v_2) & \subseteq B_2(v_1) \Rightarrow P_2(v_2) \leq P_2(v_1), \ (0 \leq 1) \\
B_1(v_1) & \subseteq B_3(v_3) \Rightarrow P_1(v_1) \leq P_3(v_3), \ (0 \leq 1) \\
B_2(v_3) & \subseteq B_1(v_1) \Rightarrow P_2(v_3) \leq P_1(v_1), \ (0 \leq 0) \\
B_1(v_3) & \subseteq B_2(v_2) \Rightarrow P_1(v_3) \leq P_2(v_2), \ (0 \leq 1) \\
B_2(v_2) & \subseteq B_3(v_3) \Rightarrow P_2(v_2) \leq P_3(v_3), \ (0 \leq 0)
\end{align*}
\]

However, the observed choice probabilities do not satisfy the condition of theorem 4.5. The set of equations

\[
\begin{align*}
P_1(v_1) &= 0 = A_{111} \quad \text{(a)} \\
P_2(v_1) &= 0 = A_{311} + A_{313} + A_{222} + A_{223} + A_{221} \quad \text{(b)} \\
P_3(v_1) &= 1 = A_{311} + A_{313} + A_{333} + A_{333} \quad \text{(c)}
\end{align*}
\]

\(^{10}\) We owe this example to Jan Karel Lenstra
Figure 4.2: Incompatible monotonic choice probabilities ($\tilde{w}_t = (v_{2t} - v_{1t}, v_{3t} - v_{1t}), t = 1, 2, 3$).

\[
\begin{align*}
P_1(v_2) &= 0 = A_{111} + A_{311} + A_{313} + A_{213} + A_{211} & \quad (d) \\
P_2(v_2) &= 1 = A_{221} + A_{323} + A_{323} + A_{222} & \quad (e) \\
P_3(v_2) &= 0 = A_{333} & \quad (f) \\
P_1(v_3) &= 1 = A_{111} + A_{311} + A_{211} + A_{221} & \quad (g) \\
P_2(v_3) &= 0 = A_{222} & \quad (h) \\
P_3(v_1) &= 0 = A_{313} + A_{333} + A_{213} + A_{223} + A_{333} & \quad (i)
\end{align*}
\]
Discrete Choice

does not have a non-negative solution for $A$. To see this, note that equations (a), (b) and (d) imply $A_{111} = A_{211} = A_{212} = A_{222} = A_{221} = A_{333} = A_{311} = A_{313} = 0$. The values do not satisfy equation (g). □

If we compare theorems 4.3, 4.4 and 4.5, 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 $V$ becomes smaller. In theorem 4.4 we do not need (C-3) and the condition (4.30) in theorem 4.5 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 4.5. Third, if the ‘measure’ of $V$ becomes smaller the freedom in choosing the distribution of $\varepsilon$ increases. Hence, it becomes easier to satisfy the compatibility conditions. Finally, we have not considered the case that $V$ 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 4.4 hold for every bounded interval and that for every subset of $V$ the conditions of theorem 4.5 are satisfied.

4.5 An Example

In this section we apply the necessary and sufficient conditions of theorem 4.5 to choice probabilities that are generated by an NMNL model. 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

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

The stochastic components of alternatives 1 and 2 are correlated: if $\theta = 1$, then $\varepsilon_1$ and $\varepsilon_2$ are stochastically independent, if $\theta \neq 0$ the joint distribution function of $(\varepsilon_1, \varepsilon_2)$ converges to

$$F(\varepsilon_1, \varepsilon_2) = \exp \left\{ - \exp \left[ - \min(\varepsilon_1, \varepsilon_2) \right] \right\}.$$
This joint distribution implies that if e.g. \( v_1 > v_2 \), then alternative 1 is eliminated from the choice set. The parameter \( \theta \) can be interpreted as the association parameter of \( \varepsilon_1 \) and \( \varepsilon_2 \). Note that \( \varepsilon_3 \) is independent of \( \varepsilon_1 \) and \( \varepsilon_2 \).

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

\[
H_1(w) = \frac{(1 + \exp(-w_2/\theta))^{(\theta-1)}}{\exp(-w_1) + (1 + \exp(-w_2/\theta))^\theta}. \tag{4.33}
\]

The corresponding density function is

\[
h_1(w_1, w_2) = \frac{(1 + \exp(-w_1))^{(\theta-2)}}{\exp(-w_2) + (1 + \exp(-w_1/\theta))^\theta} \\
\times \left\{ \frac{2(1 + \exp(-w_1/\theta)) ^ \theta}{\exp(-w_2) + (1 + \exp(-w_1/\theta))^\theta} \right\} \frac{\theta - 1}{\theta} \\
\times \frac{1}{\exp(-w_2) + (1 + \exp(-w_1/\theta))^\theta}.
\]

(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 \( \theta \) 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 4.3. 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 (4.33), we can calculate the choice probabilities as

\[
P(z_1) = (0.36, 0.59, 0.05)^T
\]
After inspection of these choice probabilities (and figure 4.3), it is seen that they satisfy the necessary condition of corollary 4.4. Moreover, using the notation of the preceding section, we see that the choice probabilities also satisfy the necessary and sufficient condition of theorem 4.5: a non-negative solution for the $A$'s is $A_{111} = 0.13, A_{113} = 0.23, A_{211} = 0, A_{213} =$
Conclusion

0.32, \( A_{21} = 0, A_{22} = 0.02, A_{23} = 0.23, A_{32} = 0, A_{33} = 0 \) and \( A_{33} = 0.05 \). 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 4.4 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 4.3) \( B_2(v_3) \subset B_2(v_4) \), but \( P_3(v_3) > P_3(v_4) \), violating the necessary condition of corollary 4.4. 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.

4.6 Conclusion

The conditions in the theorems 4.3, 4.4, 4.5 can be used to construct rationality tests in discrete choice problems. In particular, theorem 4.5 provides a test if there is a finite number of observations, and we are reluctant to extrapolate the choice probabilities outside the set of observed non-random utility components. We can refer to tests based on theorems 4.3 and 4.4 as parametric tests, because they require the (parametric) specifications of a discrete choice model. As in the NMNL model the test then boils down to a restriction on the parameters of the model. The test of theorem 4.5 is a non-parametric test, because a specification of the discrete choice model is not needed. The example of section 4.5 shows that various parametric and non-parametric tests can lead to different conclusions. Of course, the non-parametric test is the least ambiguous, because it needs no maintained hypotheses.

For a practical test in a finite but possibly large sample we need to solve two problems. First, in a finite sample the choice probabilities are estimated with sampling variance. Hence, we must develop distribution theory to determine whether a violation of the condition is due to sampling variability. Second, to check whether a moderately large sample of choice probabilities is compatible with stochastic utility maximization we need an efficient algorithm to determine whether the equation system of theorem 4.5 has a non-negative solution.
4.A Proof of Theorem 4.1

**Theorem 4.1.** Every ARUM with $\mathcal{V} = \{v\}$ implies a random preference model $(\mathcal{R}, \mathcal{I})$. Conversely, every probability distribution $\mathcal{I}$ 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 $R_\alpha$ be an arbitrary complete and transitive strict preference ordering of the $I$ alternatives. Hence, $R_\alpha$ can be written as

$$k_1 < k_2 < k_3 < \cdots < k_I$$

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 $\varepsilon$, 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

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

with $\pi_{12}$ the probability that 1 is strictly preferred over 2, $\pi_{21} = 1 - \pi_{12}$, and $\phi$ 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 $\mathcal{I}$ over the two strict preference orderings.

Next, we consider $I = 3$. The six possible preference orderings and associated probabilities are given in table 4.A-1. An ARUM with the given non-random utility components $v$ assigns utility levels
Proof of Theorem 4.1

<table>
<thead>
<tr>
<th>Preference ordering</th>
<th>Probability</th>
<th>Event in ARUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 \succ 2 \succ 3$</td>
<td>$\pi_{123}$</td>
<td>$D_{123} : w_1 &lt; z_1, w_2 - w_1 &lt; z_2 - z_1$</td>
</tr>
<tr>
<td>$1 \succ 3 \succ 2$</td>
<td>$\pi_{132}$</td>
<td>$D_{132} : w_2 &lt; z_2, w_1 - w_2 &lt; z_1 - z_2$</td>
</tr>
<tr>
<td>$2 \succ 1 \succ 3$</td>
<td>$\pi_{213}$</td>
<td>$D_{213} : w_1 &gt; z_1, w_2 &lt; z_2$</td>
</tr>
<tr>
<td>$2 \succ 3 \succ 1$</td>
<td>$\pi_{321}$</td>
<td>$D_{321} : w_2 - w_1 &lt; z_2 - z_1, w_2 &gt; z_2$</td>
</tr>
<tr>
<td>$3 \succ 1 \succ 2$</td>
<td>$\pi_{312}$</td>
<td>$D_{312} : w_2 &gt; z_2, w_1 &lt; z_1$</td>
</tr>
<tr>
<td>$3 \succ 2 \succ 1$</td>
<td>$\pi_{321}$</td>
<td>$D_{321} : w_1 &gt; z_1, w_1 - w_2 &lt; z_1 - z_2$</td>
</tr>
</tbody>
</table>

Table 4.A-1: Preference orderings with probabilities and corresponding events in ARUM. $w_1 = \varepsilon_2 - \varepsilon_1, w_2 = \varepsilon_3 - \varepsilon_1, z_1 = v_2 - v_1, z_2 = v_3 - v_1.$

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 (4.A-1) with this joint distribution yields the random preference model with probability distribution \( H \) 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

\[
D_k = \left\{ \varepsilon \in \mathbb{R}^I \mid (\varepsilon_{k_i} - \varepsilon_1) - (\varepsilon_{k_{i+1}} - \varepsilon_1) < (v_{k_i} - v_1) \right. \\
\left. -(v_{k_{i+1}} - v_1), \quad i = 1, \ldots, I - 1 \right\} \quad k = 1, \ldots, I! \tag{4.A-2}
\]
Proof of Theorem 4.1  

Let

\[ w = \varepsilon^1 - t_{l-1}\varepsilon_1, \]

with \( \varepsilon^1 \) the \((I-1)\)-subvector of \( \varepsilon \) that does not contain the first component \( \varepsilon_1 \) and \( t_{l-1} \) an \((I-1)\)-vector of ones. Define

\[ \tilde{D}_k = \left\{ \varepsilon^1 - t_{l-1}\varepsilon_1 \in \mathbb{R}^{(I-1)} \mid \varepsilon \in D_k \right\}, \quad k = 1, \ldots, I! \]

Note that

\[ \bigcup_{k=1}^{I!} D_k = \mathbb{R}^I, \quad D_k \cap D_l = \emptyset, \quad k \neq l. \]

Hence, it is obvious that

\[ \bigcup_{k=1}^{I!} \tilde{D}_k = \mathbb{R}^{(I-1)}. \quad (4.1.3) \]

If \( w \in \tilde{D}_k \cap \tilde{D}_l \), then there are \( \varepsilon \in D_k, \tilde{\varepsilon} \in D_l \) with

\[ \varepsilon^1 - t_{l-1}\varepsilon_1 = \tilde{\varepsilon}^1 - t_{l-1}\tilde{\varepsilon}_1. \]

From equation (4.1.2) we see that \( \tilde{\varepsilon} \in D_k \), and we conclude that

\[ \tilde{D}_k \cap \tilde{D}_l = \emptyset, \quad k \neq l. \quad (4.1.4) \]

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

\[ h(w) = \pi_k \frac{\prod_{i=1}^{l-1} \phi(w_i)}{\int_{D_k} \prod_{i=1}^{l-1} \phi(s_i) ds}, \quad w \in \tilde{D}_k. \]

From equation (4.1.3) and equation (4.1.4) it follows that this is a proper density function. The joint density function of \( \varepsilon \) is:

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

\[ \square \]
4.B Proof of Theorem 4.4

Theorem 4.4. The choice probabilities \( P_i(v), i = 1, \ldots, I \) are locally compatible with stochastic utility maximization on a bounded interval \( V = [a, b] \) if and only if for all \( v \in [a, b] \) (C-1), (C-2), (C’-4) and (C-5) hold.

A modified version of this theorem has been stated by Börsch-Supan (1990). Börsch-Supan states and proves this theorem only for the sufficiency part. However, since we require \( F(\epsilon) \) to be absolutely continuous, we are able to prove that the conditions are necessary as well. In the remainder of this section, we first show that Börsch-Supan’s construction does not satisfy our definition of local compatibility, and we provide an alternative proof.

Note that we can weaken the conditions somewhat: from the proof it follows that symmetry only has to hold for \( a \leq v \leq b \). Translation invariance implies that we can replace the interval \([a, b]\) by

\[
D = \left\{ v \in \mathbb{R}^l \mid \exists \epsilon \in \mathbb{R} \text{ such that } a \leq v - \epsilon t \leq b \right\}
\]

Börsch-Supan’s proof of this theorem is not correct. His proof starts from the representation

\[
P_i(v) = \int_{-\infty}^{v^l - v^l_{I-1}} h_i(w) \, dw
\]

with \( h_i \) defined as

\[
h_i(w) = \frac{\partial^{(I-1)} H_i}{\partial w^1_1 \cdots \partial w^I_{I-1}}(w)
\]

with \( H_i(w) = H_i(v^l - v^l_{I-1}) = P_i(v) \), a function defined on \( \mathbb{R}^{(I-1)} \).

Note that \( h_i \) is not a density function because for some \( v \in \mathbb{R}^l \) we have that \( h_i(v^l - v^l_{I-1}) < 0 \). Börsch-Supan proposes a density function \( h^*_i \) such that for all \( a \leq v \leq b \)

\[
P_i(v) = \int_{-\infty}^{v^l - v^l_{I-1}} h^*_i(w) \, dw, \tag{4.B-1}
\]

that is, all observed choice probabilities are generated by the density \( h^*_i \). In equation (4.B-1), \( h^*_i(w) \geq 0 \) for all \( w \in \mathbb{R}^{(I-1)} \). However, his suggestion for \( h^*_i \) does not satisfy equation (4.B-1).
To see this, first note that if \( a \leq v \leq b \), then\(^{11}\)
\[
a^1 - b_1 v_{\ell - 1} \leq v^1 - v_1 v_{\ell - 1} \leq b^1 - a_1 v_{\ell - 1}
\] (4.B-2)

In the sequel we consider the case \( I = 3 \).

In equation \(24\), p. 382, Bösch-Supan suggests to define \( h_1^* \) by

\[
h_1^*(w) = \begin{cases} 
  h_1(w) & w_1 > a_2 - b_1, w_2 > a_3 - b_1 \\
  \frac{\partial P}{\partial w_2}(w) & w_1 = a_2 - b_1, w_2 > a_3 - b_1 \\
  \frac{\partial P}{\partial w_1}(w) & w_1 > a_2 - b_1, w_2 = a_3 - b_1 \\
  P_1(w) & w_1 = a_2 - b_1, w_2 = a_3 - b_1 \\
  0 & \text{otherwise}.
\end{cases}
\] (4.B-3)

Because \( h_1^* \) is a density function with respect to Lebesgue measure, an equivalent density function is given by

\[
h_1^{**}(w) = \begin{cases} 
  h_1(w) & w_1 > a_2 - b_1, w_2 > a_3 - b_1 \\
  0 & \text{otherwise}.
\end{cases}
\] (4.B-4)

The density functions in equations (4.B-3) and (4.B-4) are equivalent because they differ on a set of Lebesgue measure 0.

It is immediately clear that for \( a \leq v \leq b \)
\[
P_1(v) > \int_{-\infty}^{a^1 - b_1 v_{\ell - 1}} h_1^{**}(w) dw
\]

because
\[
P_1(b_1, a_2, a_3) = \int_{-\infty}^{a^1 - b_1 v_2} h_1(w) dw > 0
\]

and because \( P_1(0, v_2, a_3 - b_1) \) is increasing in \( v_2 \) for \( v_2 \geq a_2 - b_1 \) and \( P_1(0, a_2 - b_1, v_3) \) is increasing in \( v_3 \) for \( v_3 \geq a_3 - b_1 \) (\( P_1(v) \) satisfies condition (C-4) on the boundary of \([a, b]\)). We conclude that Bösch-Supan's

\(^{11}\) The interval \([c, d]\) on page 382 of Bösch-Supan's paper coincides with (4.B-2), because \( a_i - b_i = \min(a_i - b_1, b_i - a_1) \) and \( b_i - a_i = \max(a_i - b_1, b_i - a_1) \).
construction does not satisfy equation (4.B-1). The problem is that he changes the value of a density function on a set of Lebesgue measure 0.

For a correct proof we must find a density function \( h_1^* \) that satisfies equation (4.B-1). If the density function of the \( I \)-vector \( \varepsilon \) is chosen as

\[
f^*(\varepsilon) = h_1^*(\varepsilon_1 - \varepsilon_j t_{i-1}) k(\varepsilon_1)
\]

with corresponding distribution function \( F^*(\varepsilon) \), and \( k \) an arbitrary density function, then \( P_i(v) \) can be expressed as in equation (4.B-1). For all \( v \in \mathcal{V} \) and \( i = 2, \ldots, I \) the density function \( h_1^* \) must also satisfy

\[
P_i(v) = \Pr(\varepsilon_2 - \varepsilon_1 \leq \varepsilon_i - \varepsilon_1 + (v_2 - v_1) - (v_i - v_1),

\ldots, (\varepsilon_i - \varepsilon_1) \geq (v_i - v_1),

\ldots, \varepsilon_I - \varepsilon_1 \leq (v_I - v_1) - (v_i - v_1))

= \int_{v_{i-1}}^{v_i} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_1^*(w) dw_{I-1} \cdots dw_1.
\]

(4.B-5)

Hence, if we can find a density function \( h_1^* \) such that for all \( v \in \mathcal{V} \) the choice probabilities \( P_i(v) \) and \( P_1(v) \), \( i = 2, \ldots, I \) can be expressed as in equations (4.B-1) and (4.B-5), then theorem 4.4 is proved. Note that, by a change of variables and due to the symmetry condition, we can write the choice probabilities \( P_i(v) \), \( i = 2, \ldots, I \) for all \( v \in \mathbb{R}^I \) as

\[
P_i(v) = \int_{v_{i-1}}^{v_i} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_1^*(w) dw_{I-1} \cdots dw_1.
\]

(4.B-6)

However, \( h_1 \) is not a density function.

**Proof (Necessity)** Since \( F(\varepsilon) \) is by assumption absolutely continuous, we have for all \( v \in \mathcal{V} \):

\[
P_i(v) = \int_{-\infty}^{\infty} \frac{\partial F}{\partial \varepsilon_i}(\varepsilon_i - v_i, \ldots, \varepsilon_i, \ldots, \varepsilon_i - v_i + v_i) d\varepsilon_i,

i = 1, \ldots, I,
\]
Proof of Theorem 4.4

which implies that all mixed partial derivatives exist almost everywhere and can be chosen to be non-negative everywhere.

(Sufficiency) We prove the theorem for \( I = 3 \). The proofs for \( I = 4, 5, \ldots \) are notationally involved, but are completely analogous. From figure 4.B-1 we see

\[
P_I(v) = \int_{-\infty}^{\alpha_2 - b_1} \int_{-\infty}^{\alpha_3 - b_1} h_1(w_1, w_2) \, dw_2 \, dw_1
\]

\[
= \int_{-\infty}^{\alpha_2 - b_1} \int_{-\infty}^{\alpha_3 - b_1} h_1(w_1, w_2) \, dw_2 \, dw_1 +
\]

\[
+ \int_{\alpha_2 - b_1}^{\alpha_2 - b_1} \int_{\alpha_3 - b_1}^{\alpha_3 - b_1} h_1(w_1, w_2) \, dw_2 \, dw_1 +
\]

\[
+ \int_{\alpha_2 - b_1}^{\alpha_2 - b_1} \int_{\alpha_3 - b_1}^{\alpha_3 - b_1} h_1(w_1, w_2) \, dw_2 \, dw_1 +
\]

\[
+ \int_{\alpha_2 - b_1}^{\alpha_2 - b_1} \int_{\alpha_3 - b_1}^{\alpha_3 - b_1} h_1(w_1, w_2) \, dw_2 \, dw_1.
\]

(4.B-7)

We consider the four terms of equation (4.B-7) in turn. Because

\[
\int_{-\infty}^{\alpha_2 - b_1} \int_{-\infty}^{\alpha_3 - b_1} h_1(w_1, w_2) \, dw_2 \, dw_1 = P_I(0, a_2 - b_1, 0, a_3 - b_1) \geq 0,
\]

there is a density function \( g_1 \) with

\[
\int_{-\infty}^{\alpha_2 - b_1} \int_{-\infty}^{\alpha_3 - b_1} h_1(w_1, w_2) \, dw_2 \, dw_1
\]

\[
= \int_{-\infty}^{\alpha_2 - b_1} \int_{-\infty}^{\alpha_3 - b_1} g_1(w_1, w_2) \, dw_2 \, dw_1.
\]

For the second term let \( f_2 \) be an arbitrary non-negative function with

\[
\int_{-\infty}^{\alpha_3 - b_1} f_2(w_2) \, dw_2 = 1.
\]

(4.B-8)

We have

\[
\int_{\alpha_2 - b_1}^{\alpha_2 - b_1} \int_{\alpha_3 - b_1}^{\alpha_3 - b_1} h_1(w_1, w_2) \, dw_2 \, dw_1
\]
\[ = \int_{a_2}^{a_2-b_1} \int_{-\infty}^{a_3-b_1} \left\{ f_2(w_2) \int_{-\infty}^{a_3-b_1} h_1(w_1, s) \, ds \right\} \, dw_2 \, dw_1. \]

(4.8-9)

Condition \( (C') \) in theorem 4.4 implies that for \( a_2-b_1 \leq w_1 \leq b_2-a_1 \),
\[
\frac{\partial P_1}{\partial v_1}(0, w_1, a_3 - b_1) = \int_{-\infty}^{a_3-b_1} h_1(w_1, s) \, ds \geq 0.
\]

Hence, the integrand in equation (4.8-9) is non-negative. The third term is rewritten analogously. Let \( f_1 \) be an arbitrary non-negative function with
\[
\int_{-\infty}^{a_3-b_1} f_1(w_1) \, dw_1 = 1.
\]

Now we can define \( h_1^* \) for \( w_1 \leq b_2-a_1, w_2 \leq b_3-a_1 \) as
\[
h_1^*(w) = \begin{cases} 
  h_1(w) & a_2 - b_1 \leq w_1 \leq b_2 - a_1, \\
  g_1(w) & a_3 - b_1 \leq w_2 \leq b_3 - a_1,
\end{cases}
\]

From equation (4.8-6) it follows that
\[
P_2(v) = \int_{w_1}^{\infty} \int_{-\infty}^{(w_3-v_2)} h_1(w_1, w_2) \, dw_2 \, dw_1
\]
\[
= \int_{b_2-a_1}^{\infty} \int_{-\infty}^{(w_3-v_2)-(b_2-a_1)} h_1(w_1, w_2) \, dw_2 \, dw_1 +
\]

(4.8-10)
Figure 4.B-1: Integration regions for choice probabilities, $I = 3$

\[
\begin{align*}
&+ \int_{b_2-a_1}^{b_3-a_1} \int_{-\infty}^{b_1} h_1(w_1, w_2) \, dw_2 \, dw_1 + \\
&+ \int_{b_2-a_1}^{\infty} \int_{b_1+(b_3-a_1)-(b_2-a_1)}^{b_2+(a_3-a_1)} h_1(w_1, w_2) \, dw_2 \, dw_1 + \\
&+ \int_{b_2-a_1}^{b_3-a_1} \int_{b_1+(b_3-a_1)-(b_2-a_1)}^{b_2+(a_3-a_1)} h_1(w_1, w_2) \, dw_2 \, dw_1.
\end{align*}
\]
Again we consider the four terms on the right-hand side of equation (4.B-11) in turn. Because

\[
\int_{b_2-a_1}^{\infty} \int_{-\infty}^{w_1 + (a_2 - b_2) - (b_3 - a_1)} h_1(w_1, w_2) d w_2 d w_1
= P_2(0, b_2 - a_1, a_3 - b_1) \geq 0
\]

there is a density function \(g_2\) with

\[
\int_{b_2-a_1}^{\infty} \int_{-\infty}^{w_1 + (a_2 - b_2) - (b_3 - a_1)} h_1(w_1, w_2) d w_2 d w_1
= \int_{b_2-a_1}^{\infty} \int_{-\infty}^{w_1 + (a_2 - b_2) - (b_3 - a_1)} g_2(w_1, w_2) d w_2 d w_1.
\]

The second term of equation (4.B-11) can be rewritten as

\[
\int_{b_2-a_1}^{\infty} \int_{-\infty}^{w_1 + b_1} h_1(w_1, w_2) d w_2 d w_1
= \int_{b_2-a_1}^{\infty} \int_{-\infty}^{w_1 + b_1} \left\{ f_2(w_2) \int_{-\infty}^{w_1 + b_1} h_1(w_1, s) d s \right\} d w_2 d w_1
\]

(4.B-12)

with \(f_2\) as in equation (4.B-8). The integrands in the equations (4.B-12) and (4.B-9) are identical. Hence, the integrand of equation (4.B-12) is non-negative. To deal with the third term of equation (4.B-11), let \(f_3\) be a non-negative function with

\[
\int_{b_2-a_1}^{\infty} f_3(w_1) d w_1 = 1.
\]

Hence,

\[
\int_{b_2-a_1}^{\infty} \int_{w_1 + (a_2 - b_2) - (b_3 - a_1)}^{w_1 + (a_2 - b_2) - (b_3 - a_1)} h_1(w_1, w_2) d w_2 d w_1
= \int_{b_2-a_1}^{\infty} \int_{(a_2 - b_2) - (b_3 - a_1)}^{w_1 + w_2} h_1(w_1, w_1 + w_2) d w_2 d w_1
\]
By condition (C’-4) of theorem 4.4 we have for \(a_3-b_1 \leq w_2 \leq b_3-a_1\)

\[ \frac{\partial P_3}{\partial v_3}(0, b_2 - a_1, w_2) = \int_{b_2-a_1}^{\infty} h_1(s, s + w_2 - (b_2 - a_1))ds \geq 0, \]

or

\[ \int_{b_2-a_1}^{\infty} h_1(s, s + r)ds \geq 0 \]

for \((a_3-b_1)-(b_2-a_1) \leq r \leq b_3-b_2\). Hence the integrand in equation (4.B-13) is non-negative. The integrand of the final term in equation (4.B-11) is non-negative by equation (C’-4).

We can now extend the definition of \(h_1^*(w)\):

\[ h_1^*(w) = \left\{ \begin{array}{ll}
g_2(w) & w_1 > b_2 - a_1, w_2 < w_1 + (a_3 - b_1) - (b_2 - a_1) \\
f_3(w_1) & w_1 > b_2 - a_1, w_1 + (a_3 - b_1) - (b_2 - a_1) \leq w_2 \leq w_1 + (b_3 - b_2). \\
\end{array} \right. \]

By equations (4.B-9) and (4.B-12) the definition on \(h_1^*(w_1, w_2)\) on \(a_2-b_1 \leq w_1 \leq b_2-a_1, w_2 < a_3-b_1\) in equation (4.B-10) can be used for \(P_3^*(v)\) too.

Finally, we consider \(P_3^*(v)\). From figure 4.B-1 we obtain

\[ P_3^*(v) = \int_{b_3-a_1}^{\infty} \int_{-\infty}^{w_2+(b_3-a_1)} h_1(w_1, w_2)dw_1dw_2 \]
\begin{align*}
&= \int_{b_1-a_1}^{\infty} \int_{-\infty}^{w_2+(a_2-b_2)-(b_3-a_1)} h_1(w_1, w_2) \, dw_1 \, dw_2 + \\
&+ \int_{b_3-a_1}^{b_2-a_1} \int_{-\infty}^{a_2-b_2} h_1(w_1, w_2) \, dw_1 \, dw_2 + \\
&+ \int_{b_3-a_1}^{b_2-a_1} \int_{w_1+(v_3-v_2)}^{\infty} h_1(w_1, w_2) \, dw_2 \, dw_1 + \\
&+ \int_{b_3-a_1}^{b_2-a_1} \int_{w_1+(v_3-v_2)}^{b_3-a_1} h_1(w_1, w_2) \, dw_1 \, dw_2 + \\
&+ \int_{b_3-a_1}^{b_2-a_1} \int_{w_1+(v_3-v_2)}^{b_3-a_1} h_1(w_1, w_2) \, dw_2 \, dw_1. \tag{4.15}
\end{align*}

We implicitly assume that the line $w_2 = w_1 + (v_3 - v_2)$ lies below the line $w_2 = w_1 + (b_3 - b_2)$. However, the argument which follows can easily be adapted to handle the other case as well. The first term on the right-hand side of equation (4.15) is equal to $P_3(0, a_2 - b_1, b_3 - a_1)$, and we can find a density function $g_3$ such that

\begin{align*}
&\int_{b_3-a_1}^{\infty} \int_{-\infty}^{w_2+(a_2-b_2)-(b_3-a_1)} g_3(w_1, w_2) \, dw_1 \, dw_2 \\
&= P_3(0, a_2 - b_1, b_3 - a_1) \geq 0.
\end{align*}

For the second term we easily check that $h_1$ can be replaced by $h_1^*$ as defined in equation (4.10). For the third term we replace $h_1$ by $h_1^*$ as defined in equation (4.14):

\begin{align*}
&\int_{b_3-a_1}^{\infty} \int_{w_1+(v_3-v_2)}^{\infty} \left( f_3(w_1) \int_{b_3-a_1}^{w_2} h_1(s, w_2 - w_1 + s) \, ds \right) \, dw_2 \, dw_1 \\
&= \int_{b_3-a_1}^{\infty} \int_{b_3-a_1}^{b_2} h_1(w_1, w_2 + w_1) \, dw_2 \, dw_1 \\
&= \int_{b_3-a_1}^{\infty} \int_{b_3-a_1}^{b_2} h_1(w_1, w_2) \, dw_2 \, dw_1.
\end{align*}

Let $f_4$ be an arbitrary non-negative function with
\[ \int_{b_{-a_{1}}}^{\infty} f_4(w_2)dw_2 = 1. \]

Then we can replace the integrand in the fourth term of equation (4.15) by
\[ f_4(w_2) \int_{b_{-a_{1}}}^{\infty} h_1(w_1 - w_2 + s, s)ds \]
which is non-negative on the integration region. Hence, we can complete the definition of \( h_1^* \) by
\[ h_1^*(w) = \begin{cases} 
  g_3(w) & w_2 > b_3 - a_1, w_1 < w_2 + (a_2 - b_1) - (b_3 - a_1) \\
  f_4(w_2) \int_{b_{-a_{1}}}^{\infty} h_1(w_1 - w_2 + s, s)ds & w_2 > b_3 - a_1,
  w_2 + (a_2 - b_1) - (b_3 - a_1) \leq w_1 \leq w_2 + (b_2 - b_3) \end{cases} \]
Note that \( h_1^* \) integrates to 1, because \( P_1 + P_2 + P_3 = 1 \). Hence, \( h_1^* \) is a density function. It is clear that \( F^*(\varepsilon) \) can be chosen to be absolutely continuous, due to the freedom we have in choosing \( g_1(w), g_2(w) \), etc.
\[ \square \]